Looking Ahead to Patient Risk Management

We're now living in a time when sophisticated automated systems continuously produce patient test results. Yet typical QC practices are based around a batch of patient samples, or are set by default to a once daily, regulatory minimum. Take your laboratory into the era of patient risk management – with Bio-Rad as your partner.

In this article you will learn about building a QC system based around patient risk management. Related articles in the appendices provide more detail on key concepts.

Regulation is changing from "One-size-fits all QC" to doing the "Right QC"

Regulation is changing. Centers for Medicare & Medicaid Services (CMS), has recognized that "One-size-fits-all QC" is no longer appropriate due to newer technologies now available in the laboratory. What is needed is "design of an appropriate and effective QCP [QC Plan] for each laboratory and each specific test; that is the 'Right' QC!" "This new QC protocol will not necessarily reduce QC requirements, but instead, will be the 'Right' QC for [each] laboratory, its environment, patients, personnel, test systems, etc."[1] For more information on Right QC see the related article *One-size-fits-all QC vs. Right QC* (Appendix I).

Limitations of "Default" to "One-size-fits all QC"

According to the collective opinion paper on findings of the 2010 convocation of experts on laboratory quality "accurate testing and short turn around times contribute to patient care."[2]

Nonetheless, the current economic climate can force cost saving measures on laboratories that may compromise accurate testing and short turn around times. For example, the pressure to cut costs may cause default to the regulatory minimum frequency for running QC materials. This default can in turn open a window of vulnerability whenever patient results are reported between the evaluation of QC materials. For more information on the extent of the window of vulnerability see the related article *Expected Number of Patients Compromised by Failure* (Appendix II).

Since the expectation is that on average half the number of patient specimens tested between QC evaluations will be affected in the event of an undetected test system failure [3], the question becomes how often should QC materials be run? Typically, analyzer performance is verified with QC materials before running patient samples. But consider the situation of a systematic failure prior to the next QC event. Can the laboratory afford the risk of reporting patient results before QC materials have been run again and the results reviewed? In the event of an unacceptable analytical run, can the laboratory afford the delay, direct and overhead costs of repeating patient samples? Furthermore, there are some labile analytes that cannot be repeated. Surely, patient samples should be bracketed by QC materials and the QC data evaluated before reporting patient results. Perhaps then, the interval to be bracketed by running QC materials is the desired turnaround time rather than a minimum regulatory requirement.

1. DHSS, CMS, *Office of Clinical Standards and Quality/Survey and Certification Group*, Ref: S&C: 12-03-CLIA, November 4, 2011.

2. Cooper, et al, "Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality." *Clinical Chemistry Laboratory Medicine*. 2011; 49(5):793-802.

3. Parvin, C.A., Yundt-Pacheco, J., Williams, M. "Designing a quality control strategy: In the modern laboratory three questions must be answered," *ADVANCE for Administrators of the Laboratory* 2011;(5):53-54.

BIO-RAD

Laboratory Responsibility for QC

Regulations under Clinical Laboratory Improvement Amendments (CLIA) state the laboratory is responsible for having control procedures that monitor the accuracy and precision of the complete analytic process. The laboratory must establish the number, type and frequency of testing control materials. The control procedures must:

- Detect immediate errors; and,
- Monitor over time the accuracy and precision of test performance.

CLIA mandates the legal minimum frequency and number of levels:

- At least once each day patient specimens are assayed … include two control materials of different concentrations.[4]

While running two control materials of different concentrations may meet minimum regulatory requirements, it may not fully address the need to control either the measurement range or the reportable range. Two levels may not be enough to provide adequate quality control across clinical decision levels and the expected measurement range. "Concurrently using quality control samples at multiple levels, allows application of additional quality control rules that improve detection and interpretation of analytical error (i.e. proportional vs. constant, random vs. systematic)."[5] For more information on including enough levels see the related article *Concentrations of Control Materials* (Appendix III).

Again, while running QC materials just once each day may meet minimum regulatory requirements it may not be often enough to ensure that patient samples are reported with confidence. Running QC materials just once each day can open a window of vulnerability where erroneous patient results are reported before QC materials are run and QC data reviewed. For more information on the window of vulnerability see the related article *Expected Number of Patients Compromised by Failure* (Appendix II).

A Patient Based Approach to Doing the Right QC

Instead of a regulatory minimum frequency for running controls, Clinical and Laboratory Standards Institute (CLSI) expresses a patient based approach in Standard C24 "Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions; Approved Guideline." According to CLSI, "The user should determine the location of control samples within a run, keeping in mind the principle that quality control results should be evaluated before reporting patient samples from the run."

CLSI explains:

- The location of control samples should consider the type of analytical process, the kind of errors that might occur, and the protocol for reporting patient results. For example, if an analytical run corresponds to a discrete batch of samples, controls might be located at the beginning and end of the run to detect shifts, might be spaced evenly throughout the batch to monitor drift, or distributed randomly among patient samples to detect errors. In any case, the QC results would be evaluated before patient results were reported.

However:

- For a high-volume analyzer that continuously produces test results, an appropriate analytical run might be defined as a certain interval of time, then QC samples would be analyzed and evaluated at the beginning of a run and then as each run (i.e., the next time interval or defined number of samples) occurs.

Then:

- If a quality control fault is detected, results reported since the previous quality control event should be reviewed.[6]

CLIA mandates corrective action for all patient test results in the unacceptable test run. For example, when any of the following occur:

- Results of control or calibration materials, or both, fail to meet the laboratory's established criteria for acceptability.

However:

- All patient test results obtained in the unacceptable test run and since the last acceptable test run must be evaluated to determine if patient test results have been adversely affected. The laboratory must take the corrective action necessary to ensure the reporting of accurate and reliable patient test results.[7]

---

4. 42CFR493.1256
5. Clinical and Laboratory Standards Institute, *C24 Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions,* Wayne PA. Section 6.2.2 Concentrations of Analytes in Control Materials.
6. Clinical and Laboratory Standards Institute, *C24 Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions,* Wayne PA. Section 8.3 Location of Control Samples.
7. 42CFR493.1282

Laboratory Quality Control Based on Risk Management

CLSI adds to the C24 principle that quality control results should be evaluated before reporting patient samples from the run in EP23-A "Laboratory Quality Control Based on Risk Management; Approved Guideline," which expressly addresses the optimal frequency of QC material testing with respect to risk assessment. In section 5.1.6 of EP-23, CLSI states:

- The optimal frequency of performing QC sample procedures depends on built-in and other controls for a given measuring system, the stability of that measuring system, and conditions in the laboratory identified through risk assessment that could affect the reliability of testing such as staff turnover, and the clinical risk of harm to a patient if an erroneous result is reported and acted on. The frequency of control procedures should also conform to applicable regulatory and accreditation requirements.

CLSI emphasizes the advantages of testing QC materials at shorter intervals:

- Monitoring the measuring system at shorter intervals increases the likelihood that systematic errors are detected before incorrect results are reported, or decreases the time before alerting the health care provider who may have received incorrect results. For example, a laboratory that evaluates the examination (analytical) process every eight hours will identify a systematic error condition much earlier than a laboratory that monitors the examination process every 24 hours. However, the total number of specimens tested in a time interval may also influence the frequency of monitoring. For example, a laboratory that tests 2000 samples in one 24-hour period might perform control procedures several times a day, whereas a laboratory that tests 50 samples in an eight-hour shift might perform control procedures at the beginning and end of a shift. The complex relationship between the frequency of QC sample testing, frequency of false rejection, and the quality of patient results has been explored by Parvin and colleagues.[8]

In discussing EP23-A, James Nichols, PhD, chair of the subcommittee that developed the guideline, commented:

- The [CLIA] requirement for testing two levels of liquid QC every day a test is run comes from the days when labs ran just a few batches of patient samples a day. With the new, more automated analyzers, there is no longer batch analysis and patient samples are run continuously. So now, the question is, do we hold those samples until the next QC run, or do we run QC continuously, every 10, 20, or 50 samples and release results in small batches? These operational considerations lead to turn around time issues, cost issues and resource issues.[9]

Running QC continuously every 10, 20, or 50 samples and releasing results in small batches is a bracketed QC strategy.

Advantages of a Bracketed QC Strategy

As cited in EP23-A, Parvin and colleagues describe the advantages of just such a bracketed QC strategy compared to immediate release of patient results:

- Conventional QC rules originated from batch-oriented testing. There is a direct relationship between the quality of the QC specimen and the quality of the patient specimens in a batch test. The advent of discrete testing has changed this relationship. With a much higher degree of independence between tests, discrete analyzers have a weaker connection between QC specimens and patients. While the quality of a QC specimen may be acceptable, a subsequent malfunction could compromise patient specimens and remain undetected until it is identified with future QC evaluations. The number of patients that are tested between QC specimens directly affects the risk of producing a compromised result. . . . The more frequently QC specimens are evaluated the lower the average number of patients at risk for unacceptable large analytical errors due to an undetected malfunction. Logically, this implies that the way to minimize patient risk is to evaluate control specimens with each patient specimen. . . Evaluating a QC specimen with each patient will minimize patient risk but is not practical — so, what can be done? One possibility is to consider bracketing patient specimens with QC specimens and not reporting patient results until a QC specimen is successfully evaluated.

(Continued on the following page)

8. Clinical and Laboratory Standards Institute, *EP23-A Laboratory Quality Control Based on Risk Management;* Approved Guideline, Wayne PA. Section 5.1.6 Frequency of Quality Control Sample Testing.

9. Malone, B. "A New Approach to Quality Control. How Can Risk Management Help Labs?" *Clinical Laboratory News,* November 2011, Volume 37, No 11.

Bracketed QC offers several attractive design features and restores the dependency relationship between the quality of QC specimens and the quality of patient results. If evaluation of the closing QC specimen in the bracket determines that no grave, persistent error state exists, then the patient specimens were also evaluated with no grave, persistent error state.[10]

Parvin and colleagues further described the concepts of risk management and a bracketed QC strategy in a series of articles published in *ADVANCE for Administrators of the Laboratory*®:

- Learning From Laboratory Failures.
  Monitoring test system failures and QC performance can help identify opportunities for improvement.
- Statistical QC & Risk Management.
  The combination can improve the overall quality of patient results.
- A closer Look at Control Material.
  The classification of control material is an important part of the laboratory's risk profile.
- Analytical Assessment in the Clinical Laboratory.
  Assessing analytical goals when the same analyte can be tested on multiple systems is explored.
- The Focus of Laboratory Quality Control.
  Why QC strategies should be designed around the patient, not the instrument.
- Designing a Quality Control Strategy.
  In the modern laboratory three questions must be answered
- The Frequency of Quality Control Testing.
  QC testing by time or number of patient specimens and the implications for patient risk are explored.
- Recovering from an Out-Of-Control Condition.
  The laboratory must assess the impact and have a corrective action strategy.
- Sigma Metrics, Total Error Budgets, and Quality Control. Make sure your test system performance and quality control procedures are aligned with your quality goals.[11]

More detailed statistical treatment can be found in:

- Parvin, C.A., Gronowski A.N. "Effect of analytical run length on quality-control (QC) performance and the QC planning process", *Clinical Chemistry*. 1997; 43:11:2149-2154.
- Parvin, C.A., "Assessing the Impact of the Frequency of Quality Control Testing on the Quality of Reported Patient Results", *Clinical Chemistry* 2008; 54:12:2049-2054.

Laboratory Quality Control Based on Six Sigma

Collective expert opinion regarding reducing the risk of harm to the patient and QC frequency can be found in:

- Cooper, et al. "Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality", *Clinical Chemistry Laboratory Medicine*. 2011; 49(5):793-802.

Over 40 medical laboratory opinion leaders, pathologists, clinical biochemists and physicians gathered together to discuss current challenges for laboratory medicine including determining the appropriate frequency of quality control. They concluded that a risk-based approach to determining QC frequency is useful and suggested using sigma (σ) to divide tests into groups:

- >6σ (excellent tests) – evaluate with one QC per day (alternating levels between days) and a 1:3.5 s rule
- 4σ–6σ (suited for purpose) – evaluate with two levels of QC per day and the 1:2.5 s rule.
- 3σ–4σ (poor performers) – use a combination of rules with two levels of QC twice per day.
- <3σ (problems) – maximum QC, three levels, three times a day. Consider testing specimens in duplicate.

For more information on sigma (σ) see the related articles *Six Sigma Background* (Appendix IV), *Sigma Values in the Laboratory* (Appendix V), *Sigma Values calculated from the Unity™ Interlaboratory Program* (Appendix VI) and *Sigma Values and QC Strategy Design* (Appendix VII).
For more information on including enough levels see the related article *Concentrations of Control Materials* (Appendix III).

10. Yundt-Pacheco, J., Parvin C.A., "The impact of QC Frequency on Patient Results." *MLO Medical Laboratory Observer*. 2008;40(9)26-27.
11. http://laboratory-manager.advanceweb.com/Web-Extras/Online-Extras/Special-Series-Quality-Assurance.aspx, accessed on September 26, 2012.

An Integrated Approach to Doing the Right QC

The convocation of experts integrated the 6σ approach into stepwise recommendations for determining QC frequency:

- Start with the legal minimum requirements.
- Next, use a 6σ approach . . . to design an initial strategy.
- Do risk assessment to identify circumstances requiring modification of the control frequency.
- Use other factors to determine what the appropriate frequency of QC is:
  - Stability of reagents/controls/specimens/analytes
  - Quality and support of supplier
  - Expertise of staff
  - Supporting infrastructure like patient statistics
- [Under] any circumstances, there is an upper limit to the amount of time that it is acceptable to go between QC – the working group felt this limit is 24 hours when testing is being done.

Other important risk factors for determining QC frequency included:

- High-risk tests – tests where there is a large patient impact for a wrong result.
- Tests that support clinical decisions in isolation.
- Tests that do not perform well (low process capability).
- Tests that are acted upon immediately.
- Tests performed on specimens that are difficult/painful to collect (or difficult – maybe impossible to recollect).

The working group also commented on the impact of time and the number of patient samples tested:

> For the most part, patient volumes differ by the day of the week, but have consistency from week to week. Thus, time can be used as a surrogate for the number of patient samples tested. Using only the number of patients to determine QC frequency may result in problems with reagent stability if the patient volume is low. When the workflow is light, time should be used to determine the QC interval. When there is a large volume of patients, controlling batch size is the main consideration and the number of patients should be used to determine the QC interval.[12]

Returning to EP23-A "Laboratory Quality Control Based on Risk Management; Approved Guideline," CLSI reaches a similar conclusion to the convocation of experts regarding a QC strategy using QC materials. In section 5.1.1 of EP-23, CLSI states:

> The QC strategy using QC samples should include:
>
> - The frequency of QC sample test events
> - The type and number of QC samples tested per test event
> - The statistical QC limits used to evaluate the results
> - The frequency of periodic review for detecting shifts and trends
> - The actions taken when results exceed acceptable limits.

In summary, in section 5.1.1 of EP-23-A, CLSI states:

> In selecting an appropriate QC sample testing frequency, laboratories should consider that any systematic (i.e., persistent, nontransient) errors that occur after a QC sample is tested may remain undetected until the next control event. Furthermore, when an error is detected, it may not be known when the error occurred in the time interval since the last QC samples were tested. The clinical use of the test results (e.g., the impact of errors on patient care), the stability of the analyte or measurand, stability of the examination (analytical) process, the number of patient samples processed between QC events, and frequency of calibration will influence the maximum interval between control events.

12. Cooper, et al. "Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality", *Clinical Chemistry Laboratory Medicine*. 2011; 49(5):793-802.

Appendix I

One-size-fits-all QC vs. Right QC

The concept of moving from One-size-fits-all QC to Right QC was introduced in a memorandum from the Director of the Survey & Certification Group to State Survey Agency Directors dated November 4, 2011 "Initial Plans and Policy Implementation for Clinical and Laboratory Standards Institute (CLSI) Evaluation Protocol-23 (EP), Laboratory Quality Control Based on Risk Management, as Clinical Laboratory Improvement (CLIA) Quality Control (QC) Policy." "This new QC protocol will not necessarily reduce QC requirements, but instead, will be the 'right' QC for this laboratory, its environment, patients, personnel, test systems, etc." A summary of the memorandum, the memorandum itself and answers to Frequently Asked Questions for EP-23 is available at:

https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/SurveyCertificationGenInfo/Policy-and-Memos-to-States-and-Regions-Items/CMS1253857.html

Appendix II

Expected Number of Patients Compromised by Failure

With respect to the window of vulnerability that is opened whenever patient results are reported between the evaluation of QC materials, Parvin and colleagues state that:

- If we consider that a test system failure can begin at any specimen with equal probability, then the expectation is that half the number of patient specimens tested between QC evaluations will be affected in the event of an undetected test system failure.[1]

Here is the proof. We begin by assuming that failure is equally likely to occur before any patient specimen between the QC events. If there are **n** patient specimens between QC events, there are **n**+1 spaces (we count the space before the last QC event as part of this).

The expected number of patients compromised by failure – or the average number of patients compromised by failure - can be computed as the sum of the patients compromised divided by the number of possible failure locations (there are **n**+1 failure locations).

If the failure occurs prior to the first patient specimen (but after the QC), then all n patients would be compromised.

If the failure occurs prior to the second patient specimen (but after the first patient specimen), then n-1 patients would be compromised.

If the failure occurs prior to the last patient specimen (but after the second to the last patient specimen), then 1 patient would be compromised.

If the failure occurs after the last patient specimen but before the next QC, then none of the patients would be compromised.

So the sum of the patients compromised is n + n-1 + … + 1 + 0.
This is equal to the sum of the first n integers, which is equal to

$$\frac{n(n+1)}{2}$$

To compute the expected number of patients compromised by a failure, divide the sum of the first n integers by n+1,

$$\frac{n(n+1)}{2(n+1)} = \frac{n}{2}$$

or ½ the number of patients between QC events.

The expected number of patients compromised by a failure is ½ the number of patients between QC events.

1. Parvin, C.A., Yundt-Pacheco, J., Williams, M. "Designing a quality control strategy: In the modern laboratory three questions must be answered," *ADVANCE for Administrators of the Laboratory* 2011;(5):53-54.

Appendix III

Concentrations of Control Materials

In Guideline C24, CLSI states “The number of levels and concentration of quality control materials should be sufficient to determine proper method performance over the measuring range of interest. “ They continue:

For most analyte-method combinations, a minimum of two levels (concentrations) of control materials is recommended. Where possible, analyte concentrations should be at clinically relevant levels to reflect values encountered in patient specimens.

And:

Control materials may be selected to cover the measuring range. Routine testing of these materials may be helpful in confirming the expected range of the procedure.

Finally, note that:

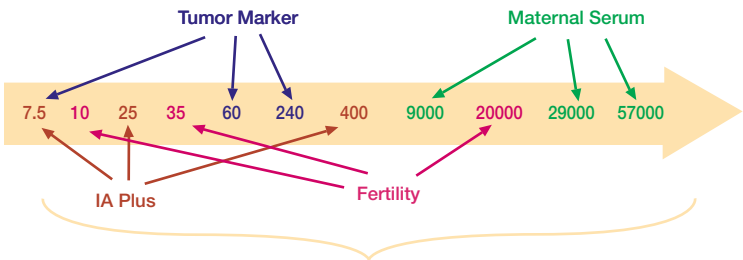
. . . the control materials specified are separate external specimens to be analyzed repeatedly by the measurement procedure, the quality control materials should be different from the calibrator materials to ensure the QC procedure provides an independent assessment of the measurement procedure’s performance in its entirety, including the procedure for calibration of the measurement.

Clinical and Laboratory Standards Institute, C24 Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions, Wayne PA. Section 6.2.1 Relation to Calibrators and Section 6.2.2 Concentrations of Analytes in Control Materials.

While for most analyte-method combinations a minimum of two levels (concentrations) of control materials is recommended, consider the analyte-method combinations that require single point calibration and have a linear detector response. Two levels of control material may not be sufficient to demonstrate linearity over the analytical measurement range (AMR). Then, consider analyte-method combinations that require multipoint calibration and have a non-linear response. Again just two levels of control material may not be sufficient to demonstrate appropriate results over the entire calibration curve. Clearly then, there are situations requiring more than a minimum of two levels (concentrations) of control materials.

For some analyte-method combinations, measuring ranges and clinical situations, adequate monitoring may have to come from more than one control material. For example, depending on the clinical situation (gestational stage; ectopic pregnancy; spontaneous abortion; trophoblastic disease) more than one control material may be needed to verify hCG performance at relevant concentrations:

Approximate Levels of hCG in Bio-Rad Controls (mIU/mL)



“Controls should verify assay performance at relevant decision points. The selection of these points may be based on clinical or analytical criteria.”

– CAP Chemistry Checklist - Jan 2012 (CHM 13900) Daily QC - Nonwalved Tests

Appendix IV

Six Sigma Background

Six Sigma is a quality management strategy that categorizes process capability by evaluating how many process standard deviations can fit within the tolerance limits of the process. The more standard deviations – or sigmas that fit between the mean of the process and the tolerance limits, the more robust the process will be.

As variation is removed from a process, the standard deviations become smaller and more of them can fit within the process tolerance limits. A process that can fit six standard deviations within the tolerance limits (a six sigma process) will only produce 3.4 defects per million opportunities even in the presence of a 1.5 standard deviation shift, and is considered "world class quality."

Sigma metrics can be computed directly from the standard deviation and the tolerance limits or they can be estimated by the defect rate. See the side bar Sigma Values in the Laboratory for details on computing laboratory sigma metrics. For estimating a sigma metric from a defect rate, determine how many defects are produced from a million opportunities. The number of defects per million opportunities can be converted into a sigma metric by comparing it to the normal distribution in terms of standard deviations.

Using the definition of the normal distribution, 0.682689 % of the distribution falls within one standard deviation, which means 0.317310 % of the distribution is expected to fall outside the one standard deviation range. Multiplying by 1,000,000, we get 317,310. If a process produces 317,310 defects per 1,000,000 opportunities, this corresponds to tolerance limits of one standard deviation, or a 1 σ process. A table can be constructed for sigma metrics 1 – 6:

Sigma Defect per 1,000,000

Sigma	Defects per 1,000,000
1σ	317,310
2σ	45,500
3σ	2,699
4σ	63
5σ	0.573
6σ	0.002

This Sigma Defects table assumes that a process is perfectly centered on the mean – usually an optimistic assumption. A more pragmatic view is taken by estimating sigma metrics assuming a 1.5 SD shift has taken place. Most sigma metric defect tables build in a 1.5 SD shift into the table, like this:

Sigma Defect per 1,000,000 with 1.5 SD Shift

Sigma	Defects per 1,000,000
1σ	691,462
2σ	308,538
3σ	66,807
4σ	6,210
5σ	2,833
6σ	3.4

Appendix V

Sigma Values in the Laboratory

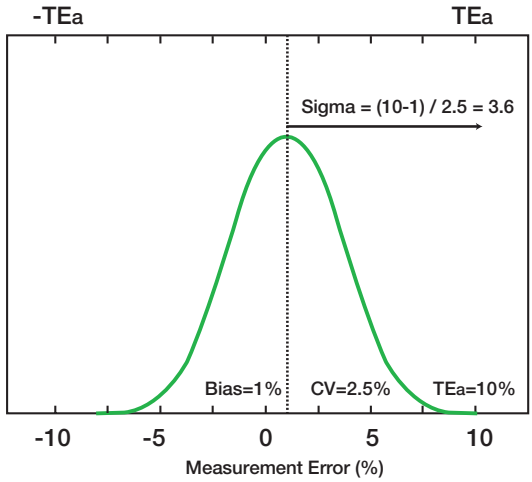
Laboratory specifications are often defined in terms of allowable total error limits (TE_a). If the difference between the true concentration of an analyte and the reported concentration in a patient's specimen exceeds TE_a the result is considered unreliable.
The sigma metric expresses the number of analytical standard deviations of the test system process that fit within the specified allowable total error limits. That is,

$$Sigma = \frac{TE_a - bias}{sd}$$

Bias is the systematic difference between the expected results obtained by the laboratory's test method and the results that would be obtained from an accepted reference. The reference may be another test method, a standard, or a consensus reference like a proficiency program or an inter-laboratory peer-comparison program. SD is the total analytical standard deviation of the test method. Equivalently, the quantities can be given as percents;

$$Sigma = \frac{\%TE_a - \%bias}{\%CV}$$

where %CV is the analytical coefficient of variation of the test method. The figure below, gives a graphical example of a test method with 1% bias, 2.5% coefficient of variation, and a specified allowable total error of 10%.



In this case the sigma value is

$$\frac{10 - 1}{2.5} = 3.6$$

That is, 3.6 analytical standard deviations fit within the 10% quality specification. Bias can have a significant impact on analytical quality and should usually be removed from the laboratory test system when it is identified. However, eliminating bias below a certain threshold can be difficult and attempts to do so are more likely to increase the overall imprecision of the test method. In general, the value for bias used in sigma computations should be the minimum threshold at which bias is actionable (i.e. an attempt to remove it will be made).

Source: "Sigma Metrics, Total Error Budgets, and Quality Control", http://laboratory-manager.advanceweb.com/Archives/Article-Archives/Sigma-Metrics-Total-Error-Budgets-QC.aspx, accessed on October 24, 2012

Appendix VI

Sigma Values calculated from the Unity™ Interlaboratory Program

Analytical imprecision was computed using QC data reported by laboratories participating in the Bio-Rad Unity™ Interlaboratory Program. 36 analytes were evaluated. 20 analytes had QC data reported at 2 concentration levels and 16 analytes had QC data reported at 3 concentration levels. QC results submitted over a 12 month period from a single lot were used to compute the mean and within-laboratory standard deviation at each concentration level of the control for every laboratory submitting.

The laboratory mean and SD at each concentration level that represented the 50th percentile (median) of the means and SDs computed for all of the individual laboratories were used in the analyses. The sigma values are representative of the QC data reported because each laboratory's sigma value was computed with respect to that laboratory's test method and then the median of all sigma values was used.

Total allowable error (TE_a) specifications were obtained from published tables that define allowable error in terms of biological variability.

When reviewing the following table recall that the Sigma Value expresses the number of analytical standard deviations that fit within the specified allowable total error limits (QC Sigma = TE_a%/CV%) so that the higher the QC Sigma the more robust the analyte and the lower the QC Sigma (σ) the less robust the analyte. For example, looking at electrolytes, potassium is a very robust analyte (5.51σ at Level 1 and 7.34σ at Level 2) whereas sodium and chloride are much less robust (1.24σ at Level 1 and 1.38σ at Level 2; and, 1.60σ at Level 1 and 1.86σ at Level 2 respectively).

The main reason there is such a difference between the sigma values for potassium, sodium and chloride is the total allowable error (TE_a) specifications used. While the CV's are relatively close, the TE_a specifications are not. Sodium has a TE_a of 1.12, Chloride 1.88, and Potassium 7.44. The large, and probably surprising differences in the performance of these electrolytes, demonstrate the utility of calculating the sigma values. The large differences in sigma values also suggest that patient results for sodium and chloride would benefit from running more Levels of QC more frequently than potassium.

See (Appendix VII) *Sigma Values and QC Strategy Design*., on the following page.

Sigma Values for Analytes with 2 QC Concentration Levels

Analytes	TE_a(%)	CV (%)		QC Sigma	
		Level 1	Level 2	Level 1	Level 2
Albumin	4.92	2.45	1.64	2.01	3.00
Alkaline Phosphatase	13.90	2.57	2.01	5.41	6.93
Amylase	17.50	1.25	1.07	14.05	16.31
Total Bilirubin	39.10	3.51	2.11	11.14	18.50
CA 125	43.80	4.06	4.06	10.79	10.80
Calcium	3.06	2.02	1.81	1.52	1.69
Chloride	1.88	1.18	1.01	1.60	1.86
Cholesterol	10.30	1.77	1.50	5.84	6.85
Creatinine	9.96	3.92	2.22	2.54	4.49
Glucose	8.88	1.68	1.43	5.27	6.23
Iron	39.70	1.97	1.00	20.16	39.83
Lactate Dehydrogenase	14.30	1.45	1.16	9.85	12.33
Lipase	37.00	2.25	2.01	16.42	18.40
Magnesium	6.03	3.07	2.50	1.96	2.41
Phosphorus	13.10	2.58	1.56	5.07	8.41
Potassium	7.44	1.35	1.01	5.51	7.34
Total Protein	4.35	1.48	1.21	2.94	3.61
Sodium	1.12	0.90	0.81	1.24	1.38
Triglycerides	35.00	2.84	1.98	12.34	17.64
Uric Acid	15.40	1.37	1.08	11.25	14.31

Turning from chemistry to immunoassay, the following table clearly identifies less robust analytes where patient results would likewise benefit from more running more Levels of QC more frequently. For example, Total T3, Free T4, Total T4 and Testosterone have Sigma Values less than 2. See related article *Sigma Values and QC Strategy Design* (Appendix VII).

Sigma Values for Analytes with 3 QC Concentration Levels

Analytes	TE_a(%)	CV (%)			QC Sigma		
		Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
CEA	29.1	7.06	4.48	3.72	4.12	6.50	7.81
Creatine Kinase	38.1	4.14	2.02	1.95	9.21	18.83	19.55
Cortisol	36.9	7.61	5.77	6.08	4.85	6.40	6.07
Ferritin	21.7	5.49	5.13	5.51	3.95	4.23	3.93
Folate	47.2	9.93	9.31	9.09	4.75	5.07	5.19
FSH	15.1	4.47	3.95	4.26	3.38	3.83	3.55
LH	24.7	4.66	4.50	5.07	5.30	5.49	4.88
Myoglobin	24.4	5.20	4.16	3.80	4.69	5.86	6.41
Prolactin	23.4	4.38	4.87	5.64	5.35	4.81	4.15
PSA	39.7	5.92	4.57	4.80	6.70	8.69	8.27
Total T3	15	8.97	6.67	5.51	1.67	2.25	2.72
Free T4	12.4	7.99	6.12	5.95	1.55	2.03	2.08
Total T4	8.7	7.36	5.35	4.75	1.18	1.63	1.83
Testosterone	17.2	10.38	9.42	9.11	1.66	1.83	1.89
TSH	29.4	6.01	4.50	4.67	4.89	6.54	6.29
Vitamin B12	16.1	6.11	6.50	10.54	2.64	2.48	1.53

Adapted from "Computing a Patient-Based Sigma Metric" Kuchipudi, L., Yundt-Pacheco, J., Parvin, C.A., *Clinical Chemistry*. 2010, 56(6), Supplement:A35.

A table of showing values for Imprecision, Bias and Total Error can be found at http://www.qcnet.com/Portals/0/PDFs/BVValues1Final.pdf

The values are derived from: Ricos, C., Alvarez, V., Cava, F., Garcia-Lario, J.V., Hernandez, A., Jimenez, C.V., Mininchela, J., Perich, C., Simon, M., "Current databases on biologic variation: pros, cons and progress", *Scandinavian Journal of Clinical and Laboratory Investigation,* 1999;59:491-500.
These values are updated/modified with the most recent specifications made available in 2012.

Appendix VII

Sigma Values and QC Strategy Design

Sigma values are useful for guiding QC strategy design. For a high sigma process it is relatively easy for the laboratory to design a QC procedure to detect any out-of-control condition that could pose a significant risk of producing unreliable results. The reason this is true is because a relatively large out-of-control condition would have to occur before there would be much chance of producing results that contained errors that exceed the allowable total error specification and it is easy to design QC procedures that can detect large out-of-control conditions. On the other hand, for a low sigma process a relatively small out-of-control condition may pose an unacceptably high risk of producing unreliable patient results. It can be challenging to design QC produces that are good at detecting small out-of-control conditions.

How Many QC Samples Should be Run?

Simple guidelines for choosing the number of QC samples to run and appropriate quality control rules based on sigma values have been proposed (Westgard JO. Six sigma quality design & control, 2nd ed. Madison WI: Westgard QC Inc., 2006). An example of one such guideline is shown in the table below:

Sigma Value	Number of QC Samples to Run	QC Rule Applied
≥6	2	$1:3_S$ or greater
5	2 or 3	$1:2.5_S$ or $1:3_S$
4	3 or 4	Multirule
3.5	6	Multirule
<3.5	Max affordable	Multirule

For lower sigma values more QC samples and more powerful QC rules are recommended. Note, a 1:3S QC rule rejects if any of the QC results differ from their target concentration by more than 3 standard deviations. Mulitrules are combinations of individual QC rules that tend to be more powerful than simple rules such as the 1:3S QC rule. In general, for large sigma value processes (≥6σ) simple QC rules with low false rejection rates are adequate. For intermediate sigma value processes (sigma values between 3.5 and 6) quality goals can be met, but more elaborate QC strategies may be required. For low sigma values (<3.5 sigma) it will be difficult to meet the laboratory's quality goals without finding ways to further reduce the test systems analytical bias, or its analytical imprecision.

How Often Should QC Samples be Run?

As well being used to guide as how many QC samples to run and the QC rules to be applied to the result, the sigma metric has also been proposed as the starting point for deciding how often controls should be run:

- >6σ (excellent tests) – evaluate with one QC per day (alternating levels between days) and a 1:3.5 s rule
- 4σ–6σ (suited for purpose) – evaluate with two levels of QC per day and the 1:2.5 s rule
- 3σ–4σ (poor performers) – use a combination of rules with two levels of QC twice per day
- <3σ (problems) – maximum QC, three levels, three times a day. Consider testing specimens in duplicate

Cooper, et al. "Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality", *Clinical Chemistry Laboratory Medicine*. 2011; 49(5):793-802.

The basic methodology is to group analytes by their sigma metric and then make frequency decisions based on patient volume and risk assessment with the low sigma analytes controlled the most frequently and high sigma analytes controlled the least frequently. See the related articles in this booklet, *Sigma Values calculated from the Unity™ Interlaboratory Program* (Appendix VI) and *Expected Number of Patients Compromised by Failure* (Appendix II).