Basic & Intermediate Quality Control Systems

Enhancing the Knowledge and Performance in Your Laboratory
Table of Contents

What is Quality Control? ................................................................. 2

How Often Should Controls be Run? ........................................... 2

Basic Quality Control Statistics ...................................................... 3
  Mean [x] .................................................................................. 3
  Standard Deviation [s] ............................................................. 3
  Coefficient of Variation [CVR] ................................................... 5
  Standard Deviation Index [SDI] ................................................... 6

Using Westgard Rules ................................................................. 6

Optional Protocols and Statistical Challenges .............................. 10
  Standard Deviation Index [SDI] ................................................ 10
  Cumulative Sums [CUSUM] ..................................................... 10
  Anion Gap ............................................................................ 11
  Patient Precision Testing ......................................................... 12

Written by
Greg Cooper, CLS, MHA
Manager of Clinical Standards and Practices

© 2009 Bio-Rad Laboratories, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of Bio-Rad Laboratories, Inc.
What is Quality Control?

Quality control in the medical laboratory is a statistical process used to monitor and evaluate the analytical process that produces patient results.

Requirements for the Statistical Process

<table>
<thead>
<tr>
<th>JULY</th>
<th>Regular testing of quality control products along with patient samples.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comparison of quality control results to specific statistical limits (ranges).</td>
</tr>
</tbody>
</table>

When a diagnostic test is performed in the medical laboratory, the outcome of the test is a result. The result may be a patient result or it may be a quality control (QC) result. The result may be quantitative (a number) or qualitative (positive or negative) or semi-quantitative (limited to a few different values).\(^1\)

QC results are used to validate whether the instrument is operating within pre-defined specifications, inferring that patient test results are reliable. Once the test system is validated, patient results can then be used for diagnosis, prognosis, or treatment planning. For example, when a patient’s serum is assayed (tested) for potassium, the test result tells us how much potassium (concentration) is present in the blood. This result is then used by the physician to determine whether the patient has a low, normal or high potassium.

Let’s assume the measured value of potassium in a patient’s serum is 2.8 mmol/L (a unit of measure, millimoles per liter).\(^2\) This result is abnormally low and indicates an inappropriate loss of potassium. But how does the person performing the test know that this result is truly reliable? It could be possible that the instrument is out of calibration and the patient’s true potassium value is 4.2 mmol/L – a normal result. The question of reliability for most testing can be resolved by regular use of quality control materials and statistical process control.

How Often Should Controls be Run?

Ideally, controls should be assayed with each analytical run and placed randomly through the run to detect analytical imprecision. Controls should also have assay values within clinically significant ranges.

Factors to Help Determine QC Frequency

- Instrument, reagent, and method reliability
- The clinical application of the test result (i.e. will incorrect patient results pose risk?)
- Amount of time available to retrieve and correct a result in error (i.e. will result be acted upon immediately after it’s reported)
- Training and competency of test operators

Use of multiple levels of control allows for better laboratory decisions regarding analytical error and validity of the run.\(^3\)

---

1. This booklet will deal only with the quality control of quantitative data.
2. Potassium can also be measured in milliequivalents per liter (mEq/L).
3. For example,
Basic Quality Control Statistics

The expected range of values for a control is calculated by use of relatively simple statistics. These statistics include mean, standard deviation, coefficient of variation, and the standard deviation index.

Mean \([\bar{x}]\)

The mean is defined as the arithmetic average of a set of data points. It is expressed in Formula 1.

\[ \bar{x} = \frac{\sum x_n}{n} \]

Where:
- \(\Sigma = \text{sum}\)
- \(x_n = \text{each value in the data set}\)
- \(n = \text{the number of values in the data set}\)

The mean describes the “central tendency” of the data set. In the clinical lab, the mean identifies the “target value” of a set of data points, usually QC or patient data.

The mean is the fundamental statistic used for comparison or for calculation of other statistics. The Clinical and Laboratory Standards Institute’s Guidance for statistical quality control recommends that at least 20 data points collected from 20 or more “separate” runs be used to establish laboratory target values for control materials.³ Laboratories should establish their own target values using manufacturer assay values only as guides. Provisional target values may be established by running 20 replicates in less than 20 runs, but the provisional values must be replaced after data from 20 separate runs is accumulated.³ However, for purposes of this discussion, only five data points will be used in the following example.

Using LDH values, find the sum of the data \{120, 115, 110, 119, 123\}. The sum \([\sum]\) is 587 U/L. The number of values is 5 \((n = 5)\). Therefore, the mean for the LDH values is 117.4 U/L (or 587 U/L divided by 5).

Standard Deviation \([s]\)

The standard deviation \((s)\) quantifies the degree of dispersion of data points about the mean and is used to set limits upon which control result acceptability is determined. Quality control data often exhibit a “normal” or Gaussian distribution around the mean.

In a Gaussian distribution:
- 68.3% of values are within ± 1.0 standard deviation of the mean
- 95.5% of values are within ± 2.0 standard deviations of the mean
- 99.7% of values are within ± 3.0 standard deviations of the mean

Clinical and Laboratory Standards Institute, C24 Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions, Wayne PA
Formula 2: Calculating a Standard Deviation [s] For a Set of QC Values

\[ s = \sqrt{\frac{\sum (x_n - \bar{x})^2}{n - 1}} \]

Where:
- \( s \) = standard deviation
- \( \bar{x} \) = mean (average) of the QC values
- \( \sum (x_n - \bar{x})^2 \) = the sum of the squares of differences between individual QC values and the mean
- \( n \) = the number of values in the data set

To calculate the standard deviation for the set of values in the previous LDH example, begin by calculating the mean \( [\bar{x}] \):

\[ \bar{x} = \frac{(120 + 115 + 110 + 119 + 123 \text{ U/L})}{5} \]
\[ \bar{x} = \frac{587}{5} \]
\[ \bar{x} = 117.4 \text{ U/L} \]

Calculate the standard deviation \([s]\) as follows:

\[ s = \sqrt{\frac{\sum (x_n - \bar{x})^2}{n - 1}} \]
\[ s = \sqrt{\frac{(120 - 117.4)^2 + (115 - 117.4)^2 + (110 - 117.4)^2 + (119 - 117.4)^2 + (123 - 117.4)^2}{5 - 1}} \]
\[ s = \sqrt{\frac{(2.6)^2 + (-2.4)^2 + (-7.4)^2 + (1.6)^2 + (5.6)^2}{4}} \]
\[ s = \sqrt{\frac{6.76 + 5.76 + 54.76 + 2.56 + 31.36}{4}} \]
\[ s = \sqrt{\frac{101.2}{4}} \]
\[ s = \sqrt{25.3} \]
\[ s = 5.03 \text{ U/L (Rounded)} \]

Limits for data acceptability are defined using the standard deviation statistic. The range for the 1s limit would be calculated as Mean ± 1s.

Consequently, the 1s range (limit) for our LDH example is calculated as:

\[ 117.4 \text{ U/L} - 5.03 \text{ U/L} = 112.4 \text{ U/L} \]
\[ 117.4 \text{ U/L} + 5.03 \text{ U/L} = 122.4 \text{ U/L} \]

The 1s range is 112.4 to 122.4 U/L.
Approximately 68% of future data should be between 112.4 and 122.4 U/L. Approximately 32% should be less than 112.4 U/L or greater than 124.4 U/L.

**The 2s range (limit) is calculated as:**

Mean ± 2s

\[
\begin{align*}
117.4 \text{ U/L} - (2 \times 5.03 \text{ U/L}) &= 107.3 \text{ U/L} \\
117.4 \text{ U/L} + (2 \times 5.03 \text{ U/L}) &= 127.5 \text{ U/L}
\end{align*}
\]

The 2s range is 107.3 to 127.5 U/L.

Only about 4.5% of future data should be less than 107.3 U/L or greater than 127.5 U/L (i.e., only one result in 20 should be beyond these limits).

Approximately 68% of future data should be between 112.4 and 122.4 U/L. Approximately 32% should be less than 112.4 U/L or greater than 124.4 U/L.

**The 3s range (limit) is calculated as:**

Mean ± 3s

\[
\begin{align*}
117.4 \text{ U/L} - (3 \times 5.03 \text{ U/L}) &= 102.3 \text{ U/L} \\
117.4 \text{ U/L} + (3 \times 5.03 \text{ U/L}) &= 132.5 \text{ U/L}
\end{align*}
\]

The 3s range is 102.3 U/L to 132.5 U/L.

Only about 0.3% of future data should be less than 102.3 U/L or greater than 132.5 U/L. It would be very unusual to obtain a result beyond these limits.

**In the medical laboratory, these ranges (limits) are used to determine the acceptability of a test run not only on the basis of a single data point but on groups of data points as well. This topic is presented in the next section.**

Standard deviation is also valuable when comparing methods or evaluating new instruments. A method or instrument with a low standard deviation produces consistent results. The lab using an instrument or method which has high standard deviations will have less certainty about the accuracy of diagnosis or the effectiveness of treatment because of test result variability. In other words, high standard deviations (poor precision, greater variability) can affect the integrity of all results. The method or instrument selected should provide a standard deviation which is medically acceptable.4

**Coefficient of Variation [CV]**

The coefficient of variation (CV) is a measure of variability. The CV for our LDH example using the CV formula would be (5.03 U/L ÷ 117.4 U/L) (100) = 4.3%.

The CV is useful for comparisons of precision at different concentrations as long as the materials used are similar and CVs are determined under similar conditions. This statistic is commonly used to compare manufacturer claims, CAP survey results, and peer group QC reports. It can also be used as a part of the internal quality control system when performing patient precision testing.

**Formula 3: Calculating the Coefficient of Variation [CV]**

\[ CV = \left( \frac{s}{\bar{x}} \right) 100 \]

Where:

- \( s \) = standard deviation
- \( \bar{x} \) = mean

**Standard Deviation Index [SDI]**

Another statistic which is helpful to evaluate performance is the standard deviation index (SDI). This statistic, which is usually obtained by participation in an interlaboratory or proficiency testing program, is used to compare a laboratory’s results to its peer group.

The target SDI is 0.0. This would indicate that the lab’s performance is identical to the peer group average. Acceptable SDI values are between ±1.0. Any test/method/instrument which has an SDI between ±1.0 and ±1.5 may have a problem and the lab should investigate. The lab must troubleshoot and correct any test/method/instrument which has an SDI of ±2.0 or greater. The relative importance of the SDI statistic does depend, however, on the size of the peer group.

It is also useful in interpreting proficiency testing, where the laboratory’s reported result replaces the lab mean in the equation for SDI. In this case, SDI values exceeding 2 or 3 suggest a problem. The SDI statistic can be used also as part of the laboratory’s internal QC system which is presented later in this document.

**Formula 4: Calculating the Standard Deviation Index [SDI]**

\[
SDI = \frac{(\bar{x}_{Lab} - \bar{x}_{Group})}{S_{Group}}
\]

Where:
- \(\bar{x}_{Group}\) = peer group mean
- \(S_{Group}\) = standard deviation
- \(\bar{x}_{Lab}\) = laboratory mean

**Using Westgard Rules**

The elements of the Westgard system are based on the principles of statistical process control used in industry nationwide since the 1950s. There are six basic rules in the Westgard scheme. Some are designed to detect random error; others detect systematic error which may indicate a bias in the system. Westgard rules are used individually or in combination to evaluate the quality of analytical runs. Rule combinations are selected by the laboratory and should be based upon the quality required and the laboratory performance for each analytical method. The overall objective is to obtain a high probability of error detection and a low frequency of false rejection of runs.

Laboratories which use the 1_{2s} rule alone in performing their quality control will frequently reject runs which are valid. According to Westgard\(^6\), failure to allow for valid points between 2s and 3s will result in falsely rejecting the following:

- 5\% of all analytical runs when using one level of control
- 10\% of all analytical runs when using two levels of control
- 14\% of all analytical runs when using three levels of control

\(^5\) There are several laboratory QC software packages that use the Westgard scheme. Unity Real Time\(^\circ\) software from Bio-Rad Laboratories is one such package. It not only uses the basic six rules, but unlike other laboratory QC software packages, it also uses additional applications for evaluation of run quality. The Westgard Rules can be used manually in concert with Levey-Jennings charts, but manual application is less efficient.

This rule is violated when a single control observation is outside the ±2s limits. Remember that in the absence of added analytical error, about 4.5% of all quality control results will fall between the 2s and 3s limits. This rule merely warns that random error or systematic error may be present in the test system. The relationship between this value and other control results within the current and previous analytical runs must be examined. If no relationship can be found and no source of error can be identified, it must be assumed that a single control value outside the ±2s limits is an acceptable random error. Patient results can be reported.

Laboratories should never use this rule as a run rejection rule without reason. This rule may be used as a run rejection rule when the lab wishes to tightly control poor performing tests.

If a Westgard rule is violated, the technologist must review the performance of the test, consult troubleshooting guides, perhaps perform maintenance, correct any identified problems or departure from protocol, and notify the supervisor who will make decisions about reporting results and rerunning the test.

This rule identifies random error or possibly the beginning of a large systematic error. Any QC result outside ±3s violates this rule.
**RULE**

**2s**

This rule identifies systematic error only. The criteria for violation of this rule are:

- Two consecutive QC results
- Greater than 2s
- On the same side of the mean

There are two applications to this rule: within-run and across runs. The within-run application affects all control results obtained for the current analytical run. For example, if a normal (Level I) and abnormal (Level II) control are assayed in this run and both levels of control are greater than 2s on the same side of the mean, this run violates the within-run application for systematic error. If however, Level I is -1s and Level II is +2.5s (a violation of the 1s rule), the Level II result from the previous run must be examined. If Level II in the previous run was at +2.0s or greater, then the across-run application for systematic error is violated.

Violation of the within-run application indicates that systematic error is present and that it affects potentially the entire analytical curve. Violation of the across-run application indicates that only a single portion of the analytical curve is affected by the error.7

**RULE**

**R4s**

This rule identifies random error only, and is applied only within the current run. If there is at least a 4s difference between control values within a single run, the rule is violated for random error. For example, assume both Level I and Level II have been assayed within the current run. Level I is +2.8s above the mean and Level II is -1.3s below the mean. The total difference between the two control levels is greater than 4s (i.e. [+2.8s – (-1.3s)] = 4.1s).

---

7 This rule also applies to trilevel (three level) controls. Whenever any two of the three levels violate the criteria for this rule within the run, unacceptable systematic error may be present and must be resolved.
This rule detects systematic error and is applied both within and across control materials. This rule is violated within the control material when the last four control values of the same control level exceed the “same” (mean +1s) or (mean –1s) limit. The rule is violated across control materials when the last four consecutive control values for different control levels exceed the “same” (mean +1s) limit or (mean –1s) limit.

This rule detects systematic error and is applied both within and across control materials. This rule is violated within the control material when the last ten values for the same control level are all on the same side of the mean. The rule is violated across control materials when the last ten consecutive values, regardless of control level, are on the same side of the mean.
Optional Protocols & Statistical Challenges

Optional protocols and statistical challenges which can be used at the discretion of the laboratory include patient precision testing, CUSUM, calculation of anion gap and SDI.

Standard Deviation Index [SDI]

Although SDI is a statistic usually generated by participation in an interlaboratory or proficiency testing program, it can be used as a tool in monitoring internal quality control performance, as well. For example, if the laboratory suspects that a trend is occurring, this suspicion can be validated by use of the SDI statistic. In this case, rather than using peer group data in calculating the statistic, the laboratory's cumulative mean and cumulative standard deviation are used. The SDI formula is modified as follows: (Mean of suspect data points – lab cumulative mean) ÷ lab cumulative standard deviation. An SDI value greater than ±1.0 indicates a possible problem with the test.

Cumulative Sums [CUSUM]

Another technique which can be employed selectively on problematic tests to detect shifts or trends is CUSUM (Cumulative Sums).

Application of this technique is very simple. It is set up for each level of control using the established mean and standard deviation. The laboratory establishes an upper and lower threshold for each level of control. Any result beyond the thresholds triggers calculation of the CUSUM. The CUSUM calculation continues for successive results until the cumulative sum either exceeds the control limit and identifies an “out of control” situation for the test under observation, or changes sign, and the calculation is stopped.

Westgard recommends a (threshold/control) limit combination of ±(1s/2.7s) or ±(0.65s/5.1s) for more particular practitioners of this technique.° These combinations yield low frequency of false rejection of valid runs and high systematic error detection. Using the LDH example with a mean of 117 U/L and a standard deviation of 5 U/L, the CUSUM threshold/control limit combination ±(1s/2.7s) would result in a lower threshold of 112 or (mean –1s) and an upper threshold of 122 or (mean +1s). The control limit would be ±13.5 or (±/2.7s).

Each control result for the test being monitored is reviewed. If the result exceeds either threshold limit, CUSUM is initiated. The difference between the result and the threshold limit is logged along with the sign of the difference. Each subsequent control result is compared to the threshold first violated and a cumulative sum of the differences is maintained. This continues until the cumulative sum changes signs or the control limit is exceeded.

8 Westgard, J. O. et. al., Combined Shewhart - CUSUM control chart for improved quality control in Clinical Chemistry. CLIN. CHEM. 23/10, 1881–1887 (1985)
A frequently neglected tool used to spot check performance of electrolyte analyzers is calculation of the anion gap. This simple formula can be applied to electrolyte data at specific intervals (time, runs, or tests) to monitor possible analytic error. If several samples fail this simple screen during any one day or run, then all of the patient results should be reviewed for possible analytical error.

In Table 1, the first two control results do not violate either the lower or upper threshold limit. Performance is acceptable. The third control result (108) violates the lower threshold limit by −4 units. The fourth control result (123) is high but for calculation of CUSUM is compared to the lower threshold limit which was first violated. The difference is 11 units. The CUSUM becomes 7. Calculation of CUSUM ends and returns to a value of zero because the CUSUM sign changed from negative to positive. The sixth control result violates the upper threshold limit. Subsequent control results cause a violation of the upper control limit (13.5) and the test is considered to be out of control. CUSUM calculation ends when the method is declared out of control and receives corrective action.

Although no Westgard rules have been broken, this scenario demonstrates a possible bias in the system. Six of eight values are above the mean and three values almost exceed the mean +2s limit, which would have triggered the 12s warning, indicating the need to review compliance with other Westgard rules. Any corrective actions taken in response to CUSUM violation or potential violation should be noted on the table or on a separate log sheet.

### Anion Gap

A frequently neglected tool used to spot check performance of electrolyte analyzers is calculation of the anion gap. This simple formula can be applied to electrolyte data at specific intervals (time, runs, or tests) to monitor possible analytic error. If several samples fail this simple screen during any one day or run, then all of the patient results should be reviewed for possible analytical error.

### Formula 5: Anion Gap

\[
\text{Na} - (\text{CO}_2 + \text{Cl}) = 5 \text{ to } 14
\]
Patient Precision Testing

Although the purpose of controls is to validate analytical runs, they also identify potential problems with the analytical system which includes the instrument or kit, technologist, ancillary equipment and reagents. Sometimes, people working in the laboratory have difficulty deciding whether the control is at fault when an out of control situation occurs, especially if it is a repeated occurrence.

Patient precision testing is a useful mechanism to distinguish between analytical system performance and control performance. It also increases sensitivity to random errors.

Patient precision testing is relatively easy to implement. The laboratory chooses an abnormal or normal patient sample for repeat testing on the next analytical run or next day. It is a form of duplicate testing.

Limits for the maximum allowable difference between duplicate results may be derived by either of two methods. One method requires the laboratory to perform a series of replicate tests on normal and/or abnormal patient samples, defining the absolute difference between each replicate, and finally establishing an acceptable range of performance based on these replicate differences.\(^9,10\)

The other system requires the use of normal and/or abnormal patient samples and the CV for the corresponding control level. In this system, the CV is applied to the initial patient result to define an acceptable replicate range.\(^10\) When the sample is retested, the replicate value is compared to the predefined range. Calculated limits for allowable differences are applicable only over a narrow range over which the standard deviation (or CV) is fairly constant. Thus, specific concentration ranges must be defined for selecting patient specimens for use as controls.

Patient precision test results can be interpreted as follows:

1. If the control and the patient precision replicate are within acceptable limits, then there probably is no problem with the control or the system.
2. If the control and the patient precision replicate are both outside acceptable limits, then there probably is a problem with the system.
3. If the control is within acceptable limits, but the patient precision replicate is outside acceptable limits, then there is possible random error occurring or a possible problem with the integrity of the precision sample.
4. If the control is outside acceptable limits, but the patient precision replicate is within acceptable limits, then there is possible random or systematic error occurring or a possible problem with the integrity of the control.
