HORIBA Medical

Quality Control Guideline.

QCP ABX Quality Control Program

# Good Laboratory Practices for Statistical QC

Extract from: http://www.westgard.com/good-laboratory-practices-for-statistical-qc-part-ii.htm

## QC Limits and Limitations

One of the major issues identified was the need for guidelines for setting QC limits. This issue can be divided into two parts, the first dealing with the right multiple of standard deviations and the second dealing with the right estimate of the mean and standard deviation. The first part involves selecting the right control rules, e.g., 2s or 3s control limits or $1_{2s}$ or $1_{3s}$ control rules for the Levey-Jennings control chart [1-3]. The second part involves the conditions and number of measurements necessary to properly observe and estimate the mean and standard deviation of the measurement procedure or analytical method.

Many laboratories today still use 2s control limits, in spite of the fact that this practice is known to give a high level of false alarms or false rejections. To counter the problem with false alarms, laboratories try to widen the control limits by using different estimates of the method standard deviation, such as manufacturer's "bottle values," peer method SDs as observed for a group of laboratories, and sometime some form of a medically significant SD.

## Are these good or bad practices?

### CLIA Guidance

The 2003 CLIA Final Rule [4] does not prescribe how to set control limits. Instead, it provides guidance for what the laboratory and the QC procedure should achieve:

> §493.1245 Standard: Control procedures
>
> (a) For each test system, the laboratory is responsible for having control procedures that monitor the accuracy and precision of the complete analytic process.
>
> (b) The laboratory must establish the number, type, and frequency of testing control materials using, if applicable, the performance specifications verified or established by the laboratory as specified in §493.1253(b)
>
> (c) The control procedure must detect immediate errors that occur due to test system failure, adverse environmental conditions, and operator performance. Monitor over time the accuracy and precision of test performance that may be influenced by changes in test system performance and environmental conditions, and variance in operator performance.

Notice the emphasis on monitoring accuracy and precision of the complete analytic process. This can best be achieved by measuring control materials that are processed in the same manner as real patient samples. Notice that **the laboratory should take into account the performance specifications verified or established by the laboratory by method validation experiments** (that's the focus of §493.1253). That again relates to the precision and accuracy observed for the method in your own laboratory. Notice that **the objectives of QC are to detect immediate errors and monitor over time the precision and accuracy of the method**. Here again the emphasis is on the performance of the individual method in an individual laboratory.

## Principles and Assumptions of Statistical QC

Some of the fundamental principles and assumptions of statistical QC can be understood with reference to Figure 1.

Measurement procedures have an inherent variability, or random error (imprecision, precision), that can be observed by analyzing the same sample again and again. That variability can be estimated from a replication experiment and can be displayed graphically by a histogram. Alternatively, that variability can be displayed point-by-point on a control chart by plotting the value on the y-axis vs time on the y-axis. If that random variability changes, it suggests that something has changed in the measurement procedure.

Statistical QC attempts to identify such changes by comparing the currently observed variability with that observed under stable operating conditions. Control limits are drawn on the control chart to help identify conditions where the observed variability no long represents the stable performance observed earlier.

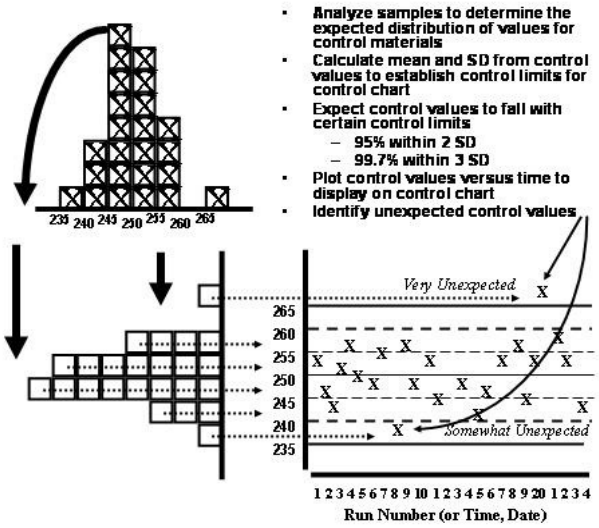


Figure 1. Fundamentals of statistical quality control: The inherent random error or imprecision of the method under stable operation is used to set control limits that will alert the analyst to conditions of unstable operation that affect the location (mean, inaccuracy) and distribution (SD, imprecision) of the method.

For statistical QC (as Internal Quality Control IQC)to work in the laboratory, it is assumed that:

- stable specimens are available which generally requires special materials developed specifically for this purpose, i.e., quality control materials;
- the variability observed is primarily due to the measurement procedure with minimum contributions from the control material itself;
- the distribution of these replicate results is assumed to be “normal” or “Gaussian,” which is reasonable for applications to a measurement procedure.

The range of variation that is expected in routine operation can be predicted from the mean and standard deviation (SD) that are calculated from the replication data.

- 95% of the results are expected to fall between the mean +2SD and the mean –2SD: This situation is also described as a 2SD control limit, i.e., a decision criterion where a run is considered out-of-control if 1 result exceeds a 2s control limit, which can also be identified as a $1_{2s}$ control rule.
- 99.7% of the results are expected to fall between the mean +3SD and the mean -3SD, which can also be described as a 3SD control limit or $1_{3s}$ control rule.

As illustrated in Figure 1, a single point that exceeds a 2SD control limit is somewhat unlikely occurrence, whereas a single point that exceeds a 3SD control limit is a very unlikely occurrence. Laboratory analysts know that 1 out of 20 or 5% of control results are expected to exceed 2SD limits, thus it is common for laboratories to just repeat the control because of the suspicion of a "false rejection."
The use of 2SD control limits can be a dangerous practice because it conditions laboratory analysts to expect false alarms, which may then lead them to ignore true alarms. When a control exceeds a 3SD limit, it is most likely a true alarm because there is such a low probability for false alarms. Ideally, QC procedures should be selected to minimize false alarms and maximize true alarms for medically important errors. It is also critical that control limits be properly established to correctly characterize the variability observed in the individual laboratory, otherwise the QC procedure will not behave as expected.

## Rejection characteristics

The behavior of different control rules (or limits) can be described by their rejection characteristics, i.e., their probabilities of false rejection and error detection [5].

- $P_{fr}$, the probability for false rejection, is the probability of a rejection occurring when there is no error except for the inherent imprecision or random variability of the measurement procedure;
- $P_{ed}$ is the probability for error detection, i.e., the probability of rejection when an error is present in addition to the inherent imprecision or random variability of the measurement procedure.

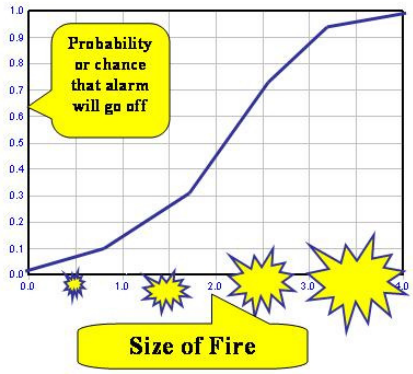


Figure 2. Typical response curve for a detector: False alarms are shown by the y-intercept of the response curve; true alarms depend on the size of the "problem". Small problems are difficult to detect, therefore the detector should be designed to detect the size of the problems of interest, e.g., errors that would be medically important.

These characteristics can be understood by analogy with a fire alarm system. You want the false alarms to be low, otherwise the alarm system itself makes us believe there are problems when there really aren't, causing us to waste time and effort. No alarm system is perfectly sensitive, thus these response curves typically are "s-shaped," starting out low, becoming steep in the middle, then leveling out at the high end, as shown in Figure 2. You would like the alarm system to be sufficiently sensitive for a problem that is important to detect, but at the same time, NOT generate any false alarms.

## Good Laboratory Practices

For statistical QC applications to behave according to principles and theory, the control limits must be properly established. This requires that both the mean and standard deviation reflect the behavior of the measurement procedure under the operating conditions in your laboratory. In other words, data from your own laboratory is necessary to characterize the mean and SD, otherwise the behavior of the QC procedure is not predictable.

For hematology tests, materials are stable for a period of a few weeks.

There is more detailed guidance from the CLSI document C24-A3 [7]:

> **C24-A3 6.2 Control Materials**: Vial-to-vial variability of the quality control material should be much less than the variation expected for the measurement procedure being monitored, and the QC materials should have demonstrated stability over their claimed shelf life, and over the claimed interval after opening the container, for the analyte of interest.

**Determine your own mean and SD.** The standard practice is to analyze 20 samples of the QC material in your own laboratory to characterize the mean and SD. The general recommendation is to obtain these 20 measurements over a 20 day period.

> **C24-A3 6.3.1 Imprecision:** Imprecision is estimated by repeated measurements on control materials during a time interval when the measurement procedure is operating in a stable condition. It is generally accepted than an initial assessment can be made by measuring a minimum of at least 20 different measurements of control material, for each control level, on separate days. Higher numbers of control measurements will provide more reliable estimates of imprecision.
>
> **C24-A3 6.3.2 Bias:** Bias should be evaluated in the context of the application of the measurement results, particularly whether they will be interpreted vs. local norms, reference limits. When interpreted vs. local norms, the focus is on the stable performance of the measurement procedure relative to a baseline event, such as a calibration. In such cases, the bias term is often assumed to be zero and the objective of statistical QC is to monitor changes from that baseline period.
>
> When results will be interpreted vs. national or international norms, measurement **bias** may be estimated…
>
> - by comparison with certified values assigned to standard reference materials…,
> - comparison of the laboratory's results with the peer group mean for external quality assessment (proficiency testing as QCP) or other interlaboratory comparison programs (EQC),
> - comparison of results obtained on a range of patient specimens analyzed by the laboratory's test method and another routine method…,

**Verify your values are within expected or labeled values.** It is good practice to utilize assayed control materials that have expected values or expected ranges. Your laboratory mean should be within the range published in the product insert. Interlaboratory Means and SDs are relevant because they reflect current testing conditions among laboratories. If the observed Means and SDs in your laboratory are not consistent with the product insert values or published interlaboratory statistics, it is very likely that your measurement procedures are not operating under the same conditions as in other laboratories. With highly automated systems, there may be accuracy or bias problems that need to be identified and fixed prior to establishing control limits, often owing to issues with calibration and standardization.

> **C24-A3 8.6.2 Assay Control Materials:** If assayed control materials are used, the values stated on the assay sheets provided by the manufacturer should be used only as guides in setting the initial control limits for testing new control materials. Actual values for the mean and standard deviation must be established by serial testing in the laboratory. The observed mean should fall within the range published by the manufacturer. EQA and peer-comparison program (QCP) provide useful measures of the means and SDs observed in other laboratories.

**Develop cumulative limits.** Obtaining 20 measurements is really a minimum for estimating the standard deviation. It would be better to have about 100 measurements, but that would take too much time to get started. The practical approach is to get 20 measurements initially, then after collecting another 20, calculate the cumulative mean, cumulative SD, and recalculate the control limits, then continue doing this periodic update until the cumulative values reflect approximately 100 measurements.

> **C24-A3 8.6.5 Cumulative Values:** Estimates of the standard deviation (and to a lesser extent the mean) from monthly control data are often subject to considerable variation from month to month, due to an insufficient number of measurements (e.g., with 20 measurements, the estimate of the standard deviation might vary up to 30% from the true standard deviation; even with 100 measurements, the estimate may vary by as much as 10%). More representative estimates can be obtained by calculating cumulative values based on control data from longer periods of time (e.g., combining control data from a consecutive six-month period to provide a cumulative estimate of the standard deviation of the measurement procedure). This cumulative value will provide a more robust representation of the effects of factors such as recalibration, reagent lot change, calibrator lot change, maintenance cycles, and environmental factors including temperature and humidity. Care should be taken to ensure that the method has been stable and the mean is not drifting consistently lower or consistently higher over the six-month periods being combined, for example due to degradation of the calibrator or control materials.

**Overlap new lot of controls.** When changing to a new lot number of control material, ideally there should be an overlap period while the new material is being analyzed to establish the new control limits. In cases where the overlap period is not sufficient, it is possible to establish the mean value for the new control material in a short time, over say a five-day period, or to start with the manufacturer's labeled mean value. Then apply the previous estimate of variation (preferably the CV) to establish the control limits. These control limits should be temporary, until sufficient data is collected to provide good estimates of both the mean and SD of the new material.

> **C24-A3 Establishing the Value of the Mean for a New Lot of QC Material:** New lots of a quality control material should be analyzed for each analyte of interest in parallel with the lot of control material in current use. Ideally, a minimum of at least 20 measurements should be made on separate days when the measurement system is known to be stable, based on QC results from existing lots. If the desired 20 data points from 20 days are not

> available, provisional values may have to be established from data collected over fewer than 20 days. Possible approaches include making no more than four control measurements per day for five different days…
>
> **C24-A3 Establishing the Value of the Standard Deviation for a New Lot of QC Material:** If there is a history of quality control data from an extended period of stable operation of the measurement procedure, the established estimate of the standard deviation can be used with the new lot of control material, as long as the new lot of material has similar target levels for the analyte of interest as for previous lots. The estimate of the standard deviation should be reevaluated periodically. If there is no history of quality control data, the standard deviation should be estimated, preferably with a minimum of 20 data points from 20 separate days… This initial standard deviation value should be replaced with a more robust estimate when data from a longer period of stable operation becomes available.

**Monitor stability of control materials with peer data.** When out-of-control problems occur, there may be concerns that the control materials themselves are causing the problems, due to deterioration over time. The best way to separate effects of your method performance from possible effects of the control materials themselves is to find out what's happening with those control materials in other laboratories. This requires access to peer data obtained on the same lot numbers of control materials. Manufacturers of control materials typically provide this information through Internet peer-comparison surveys.

> **C24-A3 9 Interlaboratory QC Programs:** When laboratories share a common pool (lot number) of control materials and report the results to an interlaboratory program, a database is created that yields statistical information, which may be used to describe or define: intralaboratory and interlaboratory imprecison; individual laboratory bias relative to a peer group; and relationship of analytical and statistical parameters of imprecision and relative bias to medical requirements. **For laboratory self-evaluation, peer-related bias and relative imprecision are useful parameters. Participation in an interlaboratory program provides an effective mechanism to complement external quality assessment** (proficiency survey) programs…

## Common or Standard Deviations from Recommended QC Practices

In the real world, there are often deviations from these standard practices. If the mean is not properly determined, the control limits will not be centered and counting rules, such $2_{2s}$, $4_{1s}$, and $10_{x}$, will be improperly triggered, giving rise to false rejections. If the SD is not properly determined, the control limits may be too wide or too narrow. If too wide, the error detection will be lost; if too narrow, false rejections will occur. There are deviant practices that occur so often, they might be considered "standard deviations" from recommended QC practices.

**Miscalculation of control limits from out-of-control results.** As additional control results are accumulated during routine operation, it is important to flag those results coming from runs that are out-of-control and to eliminate them from any future calculations of mean, SD, and control limits. This does not imply elimination from the QC records, only flagging so they are used in calculations to update the mean, SD, and control limits. Remember that the principle of statistical QC is to characterize the variation expected during stable operation, therefore only data from in-control runs should be included in the calculations. This recommendation conflicts with current practices in laboratories that use 2SD control limits, where the control ranges will narrow over time if all values outside of 2SD are eliminated. The right practice here is to eliminate the use of 2SD control limits, not try to compensate for the problems and limitations that are inherent in the use 2SD limits.

**Misuse of control range from package insert.** One common practice is to use the manufacturer's package insert values to establish control ranges, rather than data from the individual laboratory. Typically this will cause the control limits to be too wide because those values usually reflect the variation observed in several different laboratories. A too large SD will reduce the false rejections (good) but also the error detection (bad). The problem can become severe!

**Misuse of SD from a peer-comparison group.** The group SD is likely to be larger than the SD of an individual laboratory, therefore the control limits will likely be set too wide. Again, this will result in lower false rejections (good) but also lower error detection (bad). To evaluate the effect, take the ratio of the group SD to your within-lab SD and apply the multipler (2 or 3). If the group SD is twice as large as the within-lab SD and you used 2SD limits, in effect you have implemented a $1_{4s}$ rule (2*ratio SDgroup/SDwithinlab).

**Misuse of a target mean from peer-comparison group.** This seems like a reasonable practice, but it can provide misunderstanding and mismanagement of quality.

While it is okay to utilize a target mean when there is insufficient data from your own laboratory, it is critical to get your own data and switch over to your own mean as soon as possible. If the difference between your mean and the target mean from the group is large enough to be worrisome, then investigate or call the HORIBA Medical to pinpoint specific analytical problems.

**Misuse of clinical or medical control limits.** This one sounds good in theory, but is generally bad in practice. There have been some recommendations in the literature to set the control limits on the basis of clinically important changes [8], i.e., some kind of a clinical SD, rather than for statistically important changes, i.e., using the method SD. It is generally believed that the clinical SD will be larger than the statistical SD, therefore the clinical control limits will be wider than the statistical control limits. The reasoning is that a run may be out-of-control based on statistical limits, but still be okay based on clinical limits. The problem is that any control limit, however drawn, still defines a statistical control rule. To understand the true performance, you need to identify the statistical rule and assess the error detection from its power curve.

## Concluding comments

Quality practices for statistical QC mean doing the right QC right!

- The first right applies to selecting appropriate control rules and the appropriate number of control measurements to detect medically important errors, while minimizing false rejections. More detailed information is available in the scientific literature [9].
- The second right applies to implementing statistical QC properly, particularly establishing control limits correctly, as described in detail in this discussion and in the CLSI C24-A3 guideline [7].

Statistical QC is a powerful technique for managing the analytical quality of laboratory testing processes, but it must be implemented properly to provide the potential benefits. These benefits include the assurance or guarantee that analytical test results are correct for patient care and that such assurance is provided at the lowest possible cost.

## References


1. Levey S, Jennings ER. The use of control charts in the clinical laboratory. Am J Clin Pathol 1950;20:1059-66.

2. Shewhart WA. Economic Control of Quality of the Manufactured Product. New York:Van Nostrand, 1931.

3. Henry RJ, Segalove M. The running of standards in clinical chemistry and the use of the control chart. J Clin Pathol 1952;5:305-11.

4. US Centers for Medicare & Medicaid Services (CMS). Medicare, Medicaid, and CLIA Programs: Laboratory Requirements Relating to Quality Systems and Certain Personnel Qualifications. Final Rule. Fed Regist Jan 24 2003;16:3640-3714.

5. Westgard JO, Groth T, Aronsson T, Falk H, deVerdier C-H. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. Clin Chem 1977;23:1857-67.

6. Westgard JO, Barry PL. Cost-Effective Quality Control: Managing the quality and productivity of analytical processes. Washington DC:AACC Press, 1986.

7. CLSI C24-A3. Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions; Approved Guideline – Third Edition. Clinical Laboratory Standards Institute, Wayne, PA 2006.

8. Tetrault GA, Steindel SJ. Daily quality control exception practices, data analysis and critique. Q-Probes. Northfield, IL: College of American Pathologists, 1994.

9. Westgard JO. Internal quality control: Planning and implementation strategies. Ann Clin Biochem 2003;40:593-611.

10. Westgard JO. Clinical quality vs analytical performance: What are the right targets and target values? Accred Qual Assur 2004;10:10-14.