Smarter QC for stronger labs

A practical guide to modern quality control

Written by Sten Westgard

Get started >

So you want to do less QC

Quality control (QC) is a chore. No one goes home at the end of a shift and runs QC on their own time as a hobby. It's a mandatory task, one that CLIA and CMS (and a lot of other abbreviations) require of laboratories.

QC is necessary to ensure the safety of our results, but it's not sexy or exciting. However, it doesn't need to be a brick wall that we bash our head against every day.

Over more than 60 years, most laboratory medicine processes have dramatically evolved, from manual, error-prone techniques to highly automated, engineered, nearly perfected processes. The instruments from the dawn of the laboratory age look like primitive relics compared to the latest generation of diagnostic equipment.

QC, on the other hand, has barely changed in more than half a century. The same QC rules used in the 1960s are often still in use in today's laboratories.

It's time for QC to advance as far as the equipment it's supposed to monitor.

Changing your QC can be an unsettling journey where you abandon assumptions you have made and habits you've developed. In the end, though, you will have better QC, better quality, and less stress, waste, and expense.

Keep an open mind, and keep reading.

There’s the QC you have to do, and the QC you make yourself do

We can’t eliminate QC, as much as we might like to. We run QC for safety. The instruments and methods we operate are not perfect, nor are they foolproof, and we need to monitor them. We cannot outsource this responsibility to someone else—it’s our duty. We need a way to detect when things have gone wrong so we can prevent patients from being impacted by those errors.

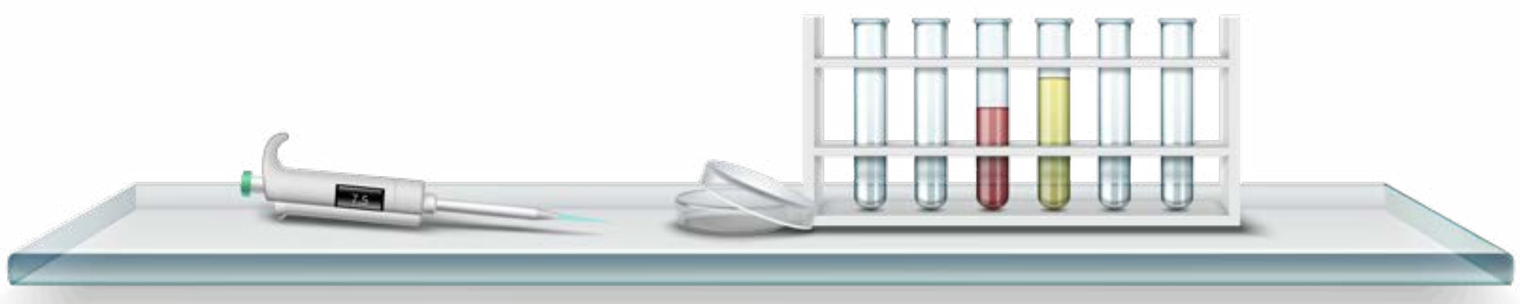
We run QC for compliance. Our regulation and accreditation guidelines are in place to enforce safety in laboratories. While we don’t always agree with how those rules are stated and measured, they ensure a minimum standard of quality in all laboratories. For most tests in the US, CLIA regulations require labs to run QC at least once a day (defining a run as a maximum of 24 hours).

Sometimes CLIA requires QC to be run more often, even two or three times a day; conversely, in extreme cases (IQCP), CLIA can be bent so a lab can reduce their QC to once a month. Running QC once every month is absurd and we don’t recommend it! Thankfully, that practice is usually restricted to point-of-care devices.

While there is a way to calculate how often you need to run QC, down to the number of patient samples that can be tested before a new control material needs to be run, we’ll explore that advanced topic later.

The frequency of QC isn’t the only way to minimize its expense; when you run QC is different from how much QC you run. The number of controls and the amount and type of control rules—all these factors can be optimized and minimized.

In other words, CLIA may mandate that your lab runs QC once a day but it doesn’t require you to use the worst QC rules and troubleshoot outliers the hardest way.

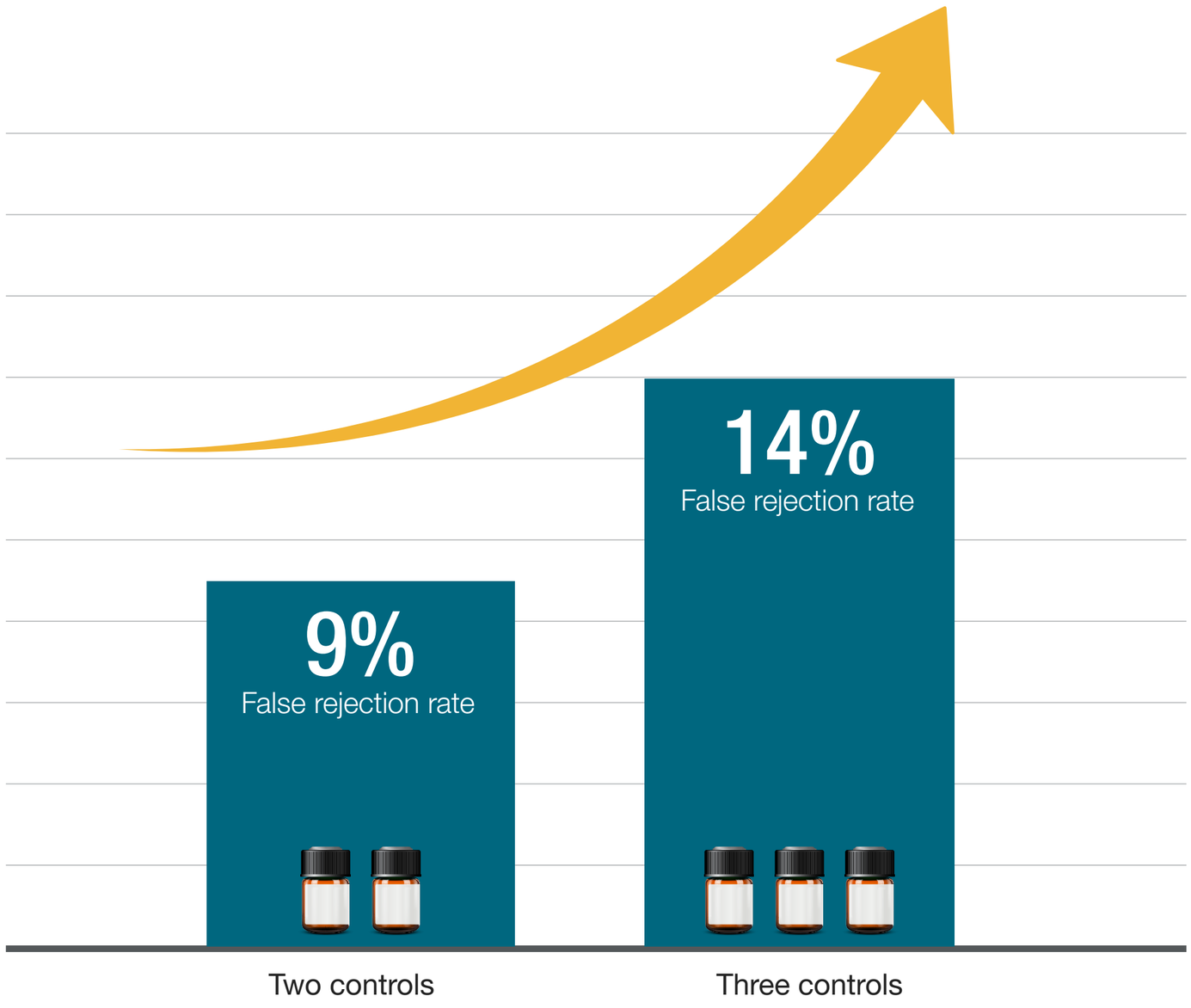


Counterproductive habits: 2 SD is too tight and too error-prone

Before there were multirule QC procedures, laboratories basically used 2 standard deviation (SD) control limits. You ran controls or (possibly) pooled patient samples and plotted them on a Levey-Jennings Shewhart chart. Anything plotted beyond 2 SD control limits was out-of-control, the run was rejected, and troubleshooting was needed.

The downside of these rules is that 2 SD limits are notoriously prone to false rejection. With 2 controls the false rejection rate is around 9%; with 3 controls, false rejections increase to around 14%. As the test menu of the early laboratories grew, false rejections proliferated. The chronic alarms generated other problems: repeating controls, running new controls, repeating those new controls, knee-jerk recalibrating, and more, meaning time and money spent for false alarms. It was clear 2 SD limits would not work with a growing test menu and expanding test volume.

If you are still using 2 SD limits in your laboratory, you're generating a lot of false rejections. **When you walk into work, you're increasing your responsibilities by 9 to 14% every day.** It's an additional burden with no added benefit.

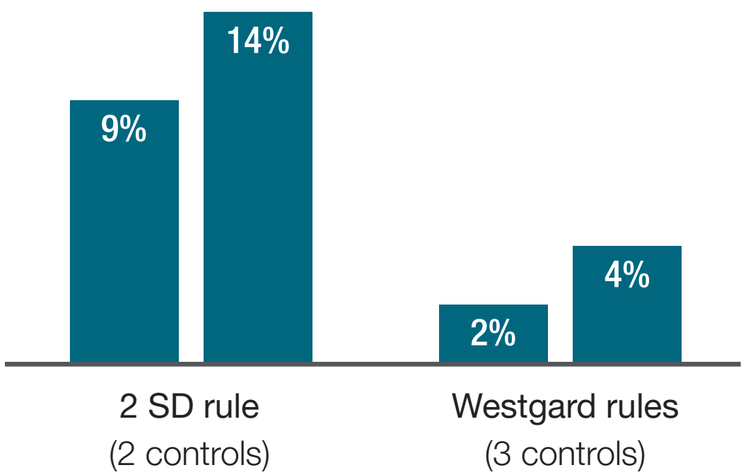


Why didn't the Westgard Rules solve this problem?

The multirule QC procedure, published in 1981, represented a breakthrough on many fronts: it was the first time a combination of QC procedures had been assessed for error detection and false rejection characteristics. It was possible to prove that the Westgard Rules could achieve high error detection without high false rejections. Instead of 9 to 14% false rejections the rejection rates were 2 to 4%, a major improvement. The replacement for 2 SD control limits had been found.

The tradition of 2 SD limits and the habit of reacting to anything past 2 SD, however, remained deeply ingrained in laboratory staff. Furthermore, the interpretation of 5 rejection rules on every control data point was a hurdle for laboratory professionals. Remember, back in the early 1980s, there were no desktop computers—QC was a manual activity carried out by humans, not software algorithms. Staff needed to be slowly weaned off 2 SD rejections while simultaneously protecting their cognitive workload. This led the 2 SD rule to be transformed from a rejection into a warning.

False rejection rates decrease with Westgard Rules



Essentially, the classic Warning Rule says, "Something might have happened, take a closer look at the data."

Lab staff would still respond to 2 SD outliers, but only to check other rules—the 5 rejection rules from the Westgard Rules. Nothing was rejected by a 2 SD outlier alone; there had to be another rule failure. This meant staff didn't have to interpret all the rules on every control point—they only had to do more work when a point jumped past the 2 SD mark—less work, less stress.

The "nudge" of 2 SD from rejection to warning, the dramatic reduction of false alarms and the carefully balanced workload of rule interpretation, made the Westgard Rules a smashing success. It also helped that they were free; not patented, not commercially exclusive, but accessible to anyone. The Westgard Rules achieved global adoption in record time—it's hard to find a QC program or an instrument that doesn't include them.

But the challenge of 2 SD outliers endures. As lab menus exploded and testing volume scaled by orders of magnitude, even the "warnings" of 2 SD began to chafe. Constant warnings meant no warnings at all—labs would become "alarm fatigued" and stop listening to the alerts, responding in wasteful ways: repeating controls automatically, often multiple times, until they got one "in" control so that they could ignore all the previous "out" controls.

Does the modern laboratory still need to be warned?

Today's laboratories don't have their lab staff draw control charts on paper, plot control points by hand, and individually interpret every point. It's done by informatics, perhaps on the instrument itself, or on an LIS, middleware, or other connected or separate software program.

Do we need to worry about overwhelming the cognitive abilities of a software algorithm? Heck, no! We can have the program interpret as many control rules as we want for each control point. There's plenty of processing capability these days. We don't need to give a warning or a heads-up to the informatics—they can handle the load.

Given this, the classic Warning Rule is obsolete, and we can just test all the control points against rejection rules. We don't have to endure false warnings. But 2 SD warnings are an enduring habit for some, part of a tradition that has been handed down from generation to generation of laboratory professionals. Because they've always used 2 SD in the past, they tell every new lab tech, "This is how we've always done QC," and the 2 SD Warning Rule persists.

Undeniably, there is psychological comfort in responding to a warning that typically turns out to be a false alarm. We see an alert but get the reassuring news that nothing's actually wrong, proving that we're vigilant. We've demonstrated that we're doing our job and we didn't have to make any real troubleshooting effort.

At its core, the classic Warning Rule is now a waste of time. When it triggers and nothing has happened, none of the rejection rules have been violated, we've put in unnecessary effort. If we just let the computer check the appropriate rejection rules immediately, we can avoid most false alarms and focus only on the real problems.

Will eliminating 2 SD limits give us enough savings?

Jettisoning the outdated practices gives us major benefits, but if we're replacing "2 SD for all" with "Westgard Rules for all," we're still missing one key modern benefit: QC doesn't need to be one-size-fits-all. Using the same rule on all tests is wasteful; some tests have reached a sixth or seventh generation of engineering, and precision once thought unachievable in the 1980s is now routine using today's methods. The best methods of the 21st century no longer need rules from the last century.

So how do we determine whether we need all the Westgard Rules or just a few of them?

The answer is Six Sigma.

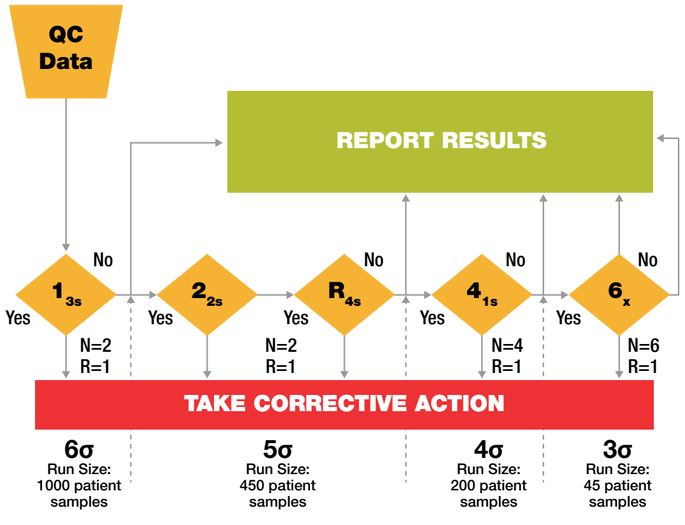
What is Six Sigma?

Maybe you've already heard of Six Sigma: it's been around for decades. There was a brief period in the 1990s to early 2000s when it was **the** quality management approach. Everyone talked about green belts, black belts, master black belts, and champions; these labels were attached to the practitioners of Six Sigma, making what would normally be considered boring quality management feel more important.

If you have one of those belts, congratulations! What we're going to discuss here doesn't require a belt, though. Six Sigma is no longer a cult, a fad, or a mysterious management technique, it's just a set of tools, like any other. YOU can do Six Sigma without expensive training, you just need to know how to count to six.

Six Sigma + Westgard Rules = Westgard Sigma Rules

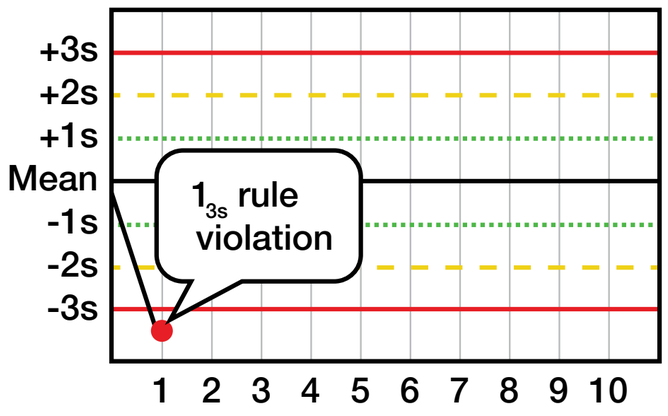
The latest version of Westgard Rules incorporates Six Sigma metrics: Westgard Sigma Rules, introduced in 2015. You're probably at least marginally familiar with Six Sigma already. You can characterize a process (in our case, an analytical method) in terms of Sigma. Six Sigma is world-class performance, representing just 3.4 defects per million (DPM) opportunities (for us, less than 4 defective test rules—either false positives or negatives—out of 1 million reportables). Five Sigma is excellent, Four Sigma is good, Three Sigma is marginal—the minimum acceptable quality—and below Three Sigma is unacceptable.



Westgard Sigma Rules greatly simplify the rules needed for different levels of quality. Not only can we calculate the analytical Sigma metric of a method, we can also connect that Sigma metric to the required amount of QC for appropriate quality.

With Six Sigma you don't need Westgard Rules, just a single rule: the **1:3s.**

$$6\sigma = 1_{3s}$$



The **1:3s** rule has a 1% false rejection rate if you're running 3 levels and less than 1% if you're running 2 levels. Compare that with the 9-14% false rejection rate that comes with the **1:2s** rule. Switching from **1:2s** to **1:3s** for Six Sigma tests dramatically reduces the number of outliers in your laboratory.

With Five Sigma, you only need a small set of rules:

$$5\sigma = 1_{3s} / 2_{2s} / R_{4s}$$

With Four Sigma, you can get good performance without the **10:x** rule:

$$4\sigma = 1_{3s} / 2_{2s} / R_{4s} / 4_{1s}$$

With Three Sigma (the minimum tolerable performance) you need the maximum Westgard Rules:

$$3\sigma = 1_{3s} / 2_{2s} / R_{4s} / 4_{1s} / 6_{x}$$

(Note that when you are running three levels of control, you need to work off the Westgard Rules adapted for three levels: **1:3s/2of3:2s/ R:4s/3:1s/9:x**).

Six Sigma

$6\sigma = 1_{3s}$

Six Sigma is 3.4 DPM, but essentially feels like perfection. While the risk is not zero, the average clinician or patient is unlikely to encounter a defect. The probability of any deviation from perfection is extremely small. Airline safety in the USA is actually better than Six Sigma, with less than 3 crashes per 1 million departures. Considering how many people fly every day, this is a very reassuring statistic. The airline industry probably wouldn't survive if it wasn't Six Sigma.

If your test is achieving Six Sigma, your clinicians and patients can feel confident in the first result you report, and probably only a single result is needed to make a diagnosis. In the laboratory, you don't need to implement extensive QC rules—in fact, you don't need rules plural, a single QC rule with wider limits will suffice. **1:3s** will provide sufficient error detection while dramatically cutting your false rejections.

Five Sigma

$5\sigma = 1_{3s} / 2_{2s} / R_{4s}$

Five Sigma is 233 DPM. Instead of single defects per million (Six Sigma), we're now in the realm of hundreds of defects. Nevertheless, on the scale of 1 million results, defects are extremely rare and will still feel "perfect" to clinicians and patients. Errors are so rare that they are almost never experienced. Results from Five Sigma methods feel just as good as Six Sigma methods.

In modern laboratories, lost test requests and lost specimens have been significantly reduced by software, automation, barcodes, cameras, and more. Considerable effort has been poured into the improvement of pre-analytical quality indicators. After more than 25 years of focus, those metrics have risen (in some cases) to Five Sigma.

Four Sigma

$4\sigma = 1_{3s} / 2_{2s} / R_{4s} / 4_{1s}$

Four Sigma is 6,230 DPM, meaning thousands of errors. Still, for over a million results, it's quite good and still a very efficient, effective process. However, enough errors are occurring that everyone is aware that mistakes can happen, and those qualms may trigger a change in behavior.

The handling of checked luggage by US airlines is Four Sigma: about 1 out of every 300 bags will end up in a different destination than the traveler. That's frequent enough that you probably know someone who won't check their luggage, thinking that 1 out of 300 is too risky. Clinicians using from Four Sigma test results may be more cautious and more likely to want to see multiple test results before making a decision.

Within the laboratory, hemolysis is often a Four Sigma process. It occurs frequently enough that it's a chronic struggle that requires sustained vigilance.

Three Sigma

$3\sigma = 1_{3s} / 2_{2s} / R_{4s} / 4_{1s} / 6_{x}$

Three Sigma is nearly 67,000 DPM: tens of thousands of errors per million results, so frequent that everyone is aware of the variation. Classic Six Sigma theory places Three Sigma as the minimum acceptable quality. It's not pretty and requires significant effort since the lab will have frequent problems and need to repeat results for clinicians who are likely to require "trending" and "baseline" parameters for a patient, given that a single result is not reliable.

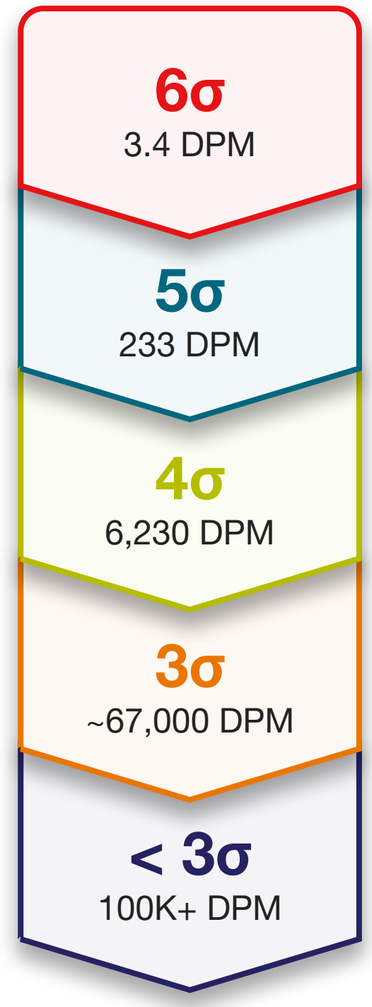
The QC failure rate of point-of-care glucose meters occurs at a Three Sigma rate, as does the contamination of microbiology slides. When you think of processes with chronic problems like this, they're probably Three Sigma.

Below Three Sigma

< 3σ

By classic Six Sigma theory, processes this error-prone are unsustainable and unreliable, requiring constant rework and extra expense and effort. Errors are occurring in the hundreds of thousands: 100,000, 200,000 or 300,000 per 1 million. The errors are constant. Results are not reliable and may be confusing for the clinician.

30% of US airline flights do not depart on time—below Three Sigma. If you think about the frustration and anger that customers associate with air travel, this is because of the subpar Sigma experience. You do not want that level of outrage associated with your laboratory results.



These are the ranges of Sigma performance and how customers experience that performance. As we saw earlier, the Sigma metric also directly determines how much QC is necessary.

Do you have to calculate the analytical Sigma Metrics, or can someone—or something—do it for you?

The best news about Sigma metrics is that they can be calculated for you. Analytical Sigma metrics calculations are a standard feature for Thermo Scientific™ LabLink™ software; you just need to select the goals and activate the calculations. There's relatively little extra work you have to perform to access Sigma metrics and optimize your QC based on them.

However, it's still useful to know how the calculation is made:

Analytical Sigma metric = [TEa - |Bias|] / SD

Two of these variables should be quite familiar to you. You already calculate your SD routinely, as you run daily controls. Bias can come from a peer comparison program (your mean versus the peer comparison group mean), a proficiency testing survey (your values versus the peer group values), or even from an assayed control (your mean versus the target mean from the package insert). The **TEa** (allowable total error) is the only variable that doesn't come from the laboratory itself. The **TEa** is the quality requirement—the goal—that is imposed by CLIA (in the US, or the RCPA in Australia, or the Rilibaek in Germany, etc.). It's a specification given to your laboratory from elsewhere.

Westgard QC provides consolidated tables of these goals, as well as recommendations for laboratories on which goals to use which are available free of charge.

Free consolidated tables:

- Consolidated Comparison of Chemistry (and Toxicology) Performance Specifications
- Consolidated Comparison of Immunoassay (and Tumor Marker) Performance Specifications
- Consolidated Comparison of Hematology and Coagulation Performance Specifications
- EFLM biological variation database

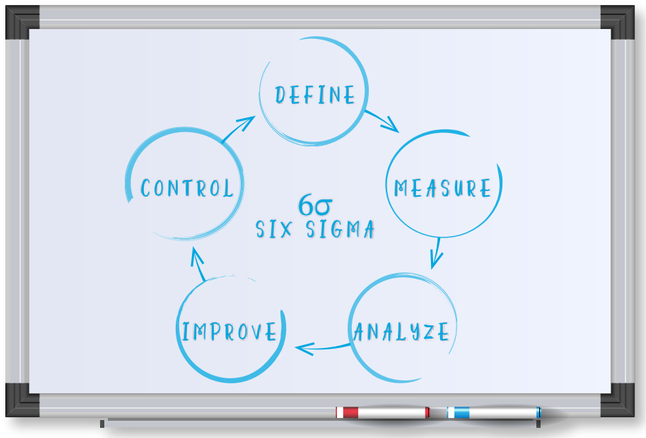
How often do you need to calculate the analytical Sigma metric?

Once the calculation is automated, it updates with every new control data point, but you don't want to track daily changes to the Sigma metric. At most, you should consider updating the metric every month. The update could coincide with your typical monthly QC review (something that many labs conduct). Other labs check their metric on a quarterly basis, others every six months. At the bare minimum, you should revisit the metric every year. Since the analytical Sigma metric is voluntary, not mandated, there are no strict rules about how often to calculate it.

Note that even if the Sigma metric changes, it might not require any change in QC. For example, all Sigma metrics above Six have the same QC recommendation, so it doesn't matter if you have Six Sigma, or even Seven, Eight, Ten, or Twenty Sigma. All these values will still only result in a **1:3s** control rule with either 2 or 3 levels of control. When the Sigma metric changes from Four to Five, or Three to Two (a full category change), then your QC rules will need to change.

In some of the more stable instruments, it's not uncommon to see a method sustain the same Sigma metric level for an extended period, months or even years at a time.

That said, there are commonsense moments when you need to start over and calculate the analytical Sigma metric anew. If significant maintenance occurs on the instrument or probes and parts are replaced, the method now has new operating characteristics, and just as you would expect to reestablish the mean and SD, you will need to recalculate the analytical Sigma metric.



What if you don't like the analytical Sigma metric you have?

There is no perfect instrument (unless you selectively populate that instrument with perfect methods and shunt all the imperfect methods to other instruments, but that's cheating). In any large menu of tests, you'll have a range of Six Sigma tests, all the way down to below Three Sigma tests.

When a test has a metric you don't like, there may be many reasons why it's not performing better, including:

1. **The problem is in the industry.** No method currently manufactured can hit the goal. The goal itself might not be appropriate, even if regulations mandate it—state-of-the-art engineering is simply unable to achieve the goal imposed (think sodium and CO2).
2. **The problem is in the manufacturer.** It is engineered to hit Three Sigma and no higher. When some instruments can hit the target but others cannot, higher performance is possible but not with the current method. It was not engineered to be able to hit the goal. When CLIA implemented new proficiency testing criteria on January 1, 2025, some methods were put into crisis. They had been engineered to hit a goal created by CLIA in 1992, and now the regulations for 2025 were 20%, 30%, or even 40% tighter.
3. **The problem is in the instrument.** Parts of the instrument are failing to live up to their design and need to be fixed: a probe, a part, or a mechanical element that is not working. A maintenance visit may turn things around.
4. **The problem is in the instrument reagents, calibrators, or other supplies.** A specific reagent lot has a bias, or a calibration is degrading, or something similar is happening to one of the "wet" parts of the instrument. These are temporary issues that may be fixed by new lots or new supplies.
5. **The problem is in the laboratory implementation.** The instrument is not being operated according to its instructions (improperly calibrated methods, poorly loaded reagents, or probes that have not been cleaned regularly are all temporary issues resulting from staff deficiencies). These problems may be addressed by stricter adherence to the SOP, more training for staff, and reinforcing the need to follow the protocols.
6. **The problem is in the QC.** The QC or the handling of the QC is introducing errors that are attributed to the instrument. "Matrix effects" might be making the analytical performance look bad, but they are merely the result of the stabilizers and preservatives and other unnatural ingredients found in the control material itself. The problems don't show up in patient samples, because none of those ingredients exist in patient samples. Alternatively, the staff are not handling the QC material correctly, causing an error that only relates to the control and doesn't have any relationship to the patient results. Better training and more commutable controls could be a solution.

Some of these problems can be fixed by the laboratory, others cannot. You can train people to do better QC handling or perform better maintenance, but you cannot get more out of a box than it was engineered to provide. You cannot hit impossible targets that no one can hit.

Even in the worst-case scenario, where there is no way to improve a low Sigma method, there are still benefits to switching to full Westgard Sigma Rules. Full Westgard Rules have two-thirds fewer false rejections than the typical **1:2s** N=2 rule applied by most labs.

For more discussion of low Sigma strategies, check out this article.

Final thoughts: Regaining control over your quality

What surveys have told us about the practice of QC is that labs are overwhelmed. They are out of control all the time. They respond to out-of-control events with self-defeating tactics: repeating, running new controls, artificially widening limits until everything falls “in range,” or ignoring QC altogether.

Amid this chaos, there is a path toward reasserted control, with rational, data-driven techniques to determine how much QC is needed for a specific test, and proven ways to reduce the wasted effort and cost of chasing constant outliers.

QC failures are often a secret shame. Labs don’t advertise their out-of-control rates. You never see a laboratory boasting of troubleshooting 10,000 QCs in a year, nor do resumes claim, “In my 10-year career as a laboratory medical professional, I have chased more than 50,000 ghosts, to no discernible benefit.”

We all shake our heads at tales of repeated repeats, or repeated-repeated repeats. It’s a ridiculous approach to QC, one that we wouldn’t accept in any other healthcare process. We wouldn’t select a surgeon who needed to repeat an operation 2, 3, or 4 times before they were successful. We wouldn’t accept a smoke detector system that went off every day even when there is no fire, but we’ve normalized this failed system. We think it’s the standard, that there is no other way.

It’s a constant grind, a daily battle, something counterproductive that saps our will repeatedly.

There is a better way, where QC rejections can be rare and also meaningful, where an out-of-control event isn’t forcing a roll of the dice, where rule violations provide meaningful information about the state of the method. Above all, there is a path where you can achieve better quality and reduce the amount of QC effort expended.

We encourage you to take the next step in this journey, and we hope we can provide you with a useful road map to this better place.

Learn more at **thermofisher.com/masqc** or **thermofisher.com/lablink360**