Six Sigma Metric Analysis for Analytical Testing Processes

Sten Westgard, M.S., Westgard QC

Introduction

Laboratories seek objective assessment and comparison of analytical methods and instrumentation performance. Unfortunately, there are few ways to compare systems on a level playing field to make an "apples to apples" comparison. Current methods of assessment can be arbitrary, relying on unclear "state of the art" assessments, or focusing more on easily tangible efficiency metrics, such as speed, cost, or ease of use. Analytical goals and requirements for the quality delivered by a test are often overlooked during the decision-making process leading to the purchase of instrumentation. Rapidly changing regulatory schemes increase the confusion over acceptable standards for instrument and method quality.

A technique to objectively and quantitatively assess the performance of methods, instruments, and laboratories is laid out in this paper. The technique consists of three components: (1) the Six Sigma metric, a widely-accepted measure of quality management, process improvement, and universal benchmarking; (2) quality requirements in the form of specific quantitative goals for analytical tests; and (3) performance data from method validation and verification studies or routine laboratory data.

One way to understand how Sigma metric analysis combines these three components is to picture a target with an arrow (Figure 1). The shape of the target is determined by Six Sigma metrics. The size of the target is determined by the size of the quality requirement. Where the arrow hits that target is determined by the method performance data.

Figure 1.

Sigma Metric Analysis provides not only an objective assessment of analytical methods and instrumentation, but it also provides the critical design information needed for operational implementation. The Sigma metric analysis process leads naturally to a quality control (QC) design scheme using quantitative and graphic tools to determine the necessary quality control procedures for routine monitoring of methods and instruments.

Adopting Six Sigma as the Goal for Laboratory Testing

Six Sigma is a widely-accepted quality management system, perhaps best known outside of healthcare as the product of innovation at General Electric and Motorola.[1] Six Sigma is also well known for the colorful titles of its practitioners – green belt (part-time Six Sigma worker), black belt (full-time Six Sigma worker), master black belt (consultant to black belts), and champion (executive proponent of Six Sigma efforts). Six Sigma has been adopted by both manufacturing and service industries, as well as healthcare institutions, from, hospitals to reference laboratories.

Six Sigma is a metric that quantifies the performance of processes as a rate of Defects-Per-Million Opportunities, (DPM, or DPMO). Six Sigma programs also encompass robust techniques such as Define-Measure-Analyze-Improve-Control (DMAIC) and Root Cause analysis to find and eliminate defects and variation within a process.

The goal of Six Sigma, in its simplest distillation, is to eliminate or reduce all variation in a process. For example, variation in a process leads to wasted effort and resources on retesting and workarounds. Reducing defects reduces costs, and improves performance and profitability. A process that achieves the goal of Six Sigma delivers both quality and efficiency.

The quantitative goal of Six Sigma is to create a process that minimizes variation until six standard deviations can fit within the tolerance limit (Figure 2). At the level of Six Sigma performance (world class quality performance), approximately three defects will occur per million opportunities.
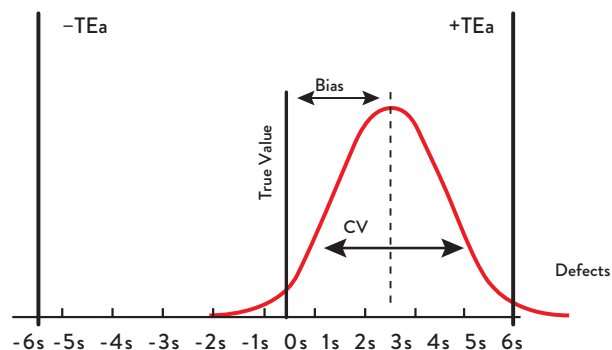

Figure 2: Relationship of imprecision (CV), inaccuracy (Bias) and allowable total error (TEa) in predicting defects

The Six Sigma scale typically runs from zero to six, but a process can actually exceed Six Sigma, if variability is sufficiently low as to decrease the defect rate. In industries outside of healthcare, 3 Sigma is considered the minimal acceptable performance for a process. When performance falls below 3 Sigma, the process is considered to be essentially unstable and unacceptable. In contrast to other industries, healthcare and clinical laboratories appear to be operating in a 2 to 3 Sigma environment. The routine use of "2s" (i.e., 2 standard deviations or 2 SD) control limits is indicative of a complacent tradition in quality control practices. Despite the well-known problems of 2s limits – they can generate false rejection rates of up to 10-20%, depending on the number of controls run – many laboratories use them for all testing processes. The misuse of 2s limits in laboratory testing frequently results in erroneously-repeated controls, excessive trouble-shooting, or worse still, workarounds that artificially widen control limits to the point that laboratories can no longer detect critical analytical errors.

Part of the power of the Six Sigma scale is its ability to provide a universal benchmark. Sigma metrics allow comparison of different processes with each other, even comparing processes across different institutions and different industries. For example, airline safety is known to be better than Six Sigma with a rate of only 1.5 crashes per million departures, while airline baggage handling in the U.S. is only 4.1 Sigma since approximately 1% of luggage is misplaced or lost, and U.S. airline departures perform at only 2.3 Sigma since nearly 30% of flights are delayed, which helps to explain chronic customer complaints.

In healthcare, the Sigma performance of common processes are less well known. When the Institute of Medicine issued its landmark report, *To Err is Human*,[2] it famously revealed that between 48,000 and 90,000 unnecessary deaths occurred in U.S. hospitals every year. Examining the death rates at the hospitals that formed the basis of the study reveals that healthcare is performing at only 3.8 Sigma. If healthcare were achieving Six Sigma, the death rate would be only 16 to 34 deaths per year

Nevalainen's[3] groundbreaking work in Sigma assessment in the clinical lab analyzed the performance of common laboratory processes and found that many were woefully inadequate Figure 3.

Q-Probe Quality Indicator	% Error	DPM	Sigma
Order accuracy	1.8%	18,000	3.60
Duplicate test orders	1.52%	15,200	3.65
Wristband errors (not banded)	.065	6,500	4.00
Therapeutic Drug Monitoring (TDM) timing errors	24.4	244,000	2.20
Hematology specimen acceptability	0.38	3,800	4.15
Chemistry specimen acceptability	0.30	3,000	4.25
Surgical pathology specimen accessioning	3.4	34,000	3.30
Cytology specimen adequacy	7.32	73,700	2.95
Laboratory proficiency testing	0.9	9,000	3.85
Surgical pathology frozen section diagnostic discordance	1.7	17,000	3.60
PAP smear rescreening false negatives	2.4	24,000	3.45
Reporting errors	0.0477	477	4.80

Figure 3: Sigma metrics of common laboratory processes, Nevalainen data.

Revisiting the arrow and target model, Six Sigma provides the shape of the target. The shape defines the goal of Six Sigma performance as the bull's-eye, as well as the inner rings of 4 and 5 Sigma. Outside the 3 Sigma ring, performance is considered to have missed the target and the process is not considered to be "fit for purpose."

Defining Quality Requirements

Knowing the shape of the target is not enough. The size must also be described. In Six Sigma terminology, the tolerance limits must be defined. In the clinical laboratory, the quality required by an analytical testing process must be defined. Tolerance limits in the laboratory are best expressed as a total allowable error (TEa) specification.

TEa is a well-accepted concept in healthcare laboratories as a model that combines both the imprecision and the inaccuracy (bias) of a method to calculate the total impact on a test result.[4] An allowable total error is the expression of how much combined imprecision and inaccuracy can be tolerated in the test result without negatively impacting patient care based on interpretation of that result.

Determining the quality required by a laboratory test is not as simple as it sounds. Most laboratories do not know the analytical quality required by their tests. Indeed, many laboratories assume that it is not even necessary to know. As long as there are no direct complaints about testing quality, many laboratories assume that the analytical

quality they are providing is adequate. This is not the only crippling assumption that laboratories make. Sometimes laboratories assume that the quality of any test is sufficient simply because a manufacturer built the instrument and made the reagents. While it's common to assume that no manufacturer would produce instruments and reagents that perform poorly, it is not good laboratory practice. Finally, laboratories frequently assume that simply following the manufacturer's directions is enough to assure the quality of the tests they provide. Again, the fact that a manufacturer provides directions does not guarantee that the directions are adequate. Professional standards as well as regulatory requirements place the burden of selecting appropriate quality control procedures on the laboratory, specifically the laboratory director.

Part of the difficulty for laboratories is defining quality specifications. Decades ago, only a few sources existed. Fortunately, a wealth of quality requirements and targets have become available and are easily obtainable. First and foremost, U.S. laboratories are governed by CLIA proficiency testing guidelines. For nearly 80 analytes, CLIA provides specific quality requirements. Other analytical benchmarks are provided by proficiency testing programs, external quality assurance programs, or peer groups. Outside the U.S., some quality specifications are available from the Royal College of Australasian Pathologists (RCPA), as well as the Guidelines of the German Medical Association (RiliBÄk).

Clinical benchmarks can also be used to generate quality requirements. Dr. Carmen Ricos and her colleagues have provided a continuously updated database of biologic variation since 2000. For more than 300 different analytes, they have tabulated desirable specifications for imprecision, inaccuracy and total allowable error.[5] ISO 15189, the new international lab accreditation standard for quality in laboratories, also provides guidance on analytical testing. Finally, the growing body of research on Evidence-Based Laboratory Medicine (EBLM) guidelines can be used to develop clinical decision intervals. These intervals can, in turn, be used to determine quality requirements for individual tests. At the very least, a laboratory can consult the clinicians who use its test results and, by documenting how test results are interpreted, determine the quality required by their testing processes.

Returning to the arrow and target model once more, establishing quality requirements, determines the size of the target. Since the use and performance of different tests varies, so too does the size of the target that the arrow/process must hit. Together, Six Sigma and quality requirements provide the shape and size of the target. Now all that remains is determining where (and if) the arrow hits the target. For that, we need data on the actual performance of the process.

Measuring Six Sigma Performance in the Laboratory

Usually, Sigma performance is assessed by counting defects, then converting that count into a Defects Per Million Opportunities (DPM, or DPMO) rate. Once the DPM is known, a Six Sigma table, available in standard text books, can be consulted to obtain the Six Sigma metric.

Counting defects relies on two capabilities. First, it must be possible to define what a process defect means. Second, it must be possible to detect a process defect when it occurs. For most processes, these are simple tasks. Most processes analyzed in Six Sigma projects use the counting defect approach.

In the laboratory, counting defects is also the usual Six Sigma metric technique. For example, turn around time (TAT) is very easy to define for laboratories. A laboratory might set a target of returning test results within 60 minutes of specimen receipt. Thus, when a test result is returned after 61 minutes, it's simple to detect the defect (i.e., TAT is > 60 minutes). Counting the number of defective test results (> 60 minutes) over a period of time is an easy way to determine the Sigma performance of the laboratory's TAT.

For laboratory test results, however, determining and detecting defects is more difficult. When a single test result is generated, it's not possible to know what the true value of that test result should be, even if the sample is tested multiple times. For example, if a cholesterol test result is 212 mg/dL, the "true value" of that test is not known, unless the specimen was also analyzed by an accepted reference method. Thus, it's unknown if the result falls

within the tolerance limits or quality requirements. If the true value is 190 mg/dL, the observed test result is probably a defect. If the true value is 205 mg/dL, the observed test result is probably acceptable. But without knowing the true value, there is no way of counting how many defects are being generated by a testing process.

Fortunately, there is another method of determining the Sigma metric of a process: by measuring variation. Conveniently for laboratories, measuring variation through the use of controls is part of the daily routine. Controls are a known value, so variation of an observed test result can be measured. With multiple control results, information on the standard deviation of testing processes can be collected and the imprecision (coefficient of variation, % CV) can be calculated. Information about the inaccuracy (bias) of an analytical testing process can readily be calculated by comparing results between the testing method and a reference method, or by analyzing the results of the testing method in proficiency testing, peer group, or some other form of external quality assurance program.

Ideally, the data on imprecision and inaccuracy is collected during the same time frame and at the same critical level (medical decision level) of test interpretation. In other words, the data on performance should be an accurate snapshot of method performance at a specific point in time and at a specific concentration of analyte. Thus, the resulting Sigma metric best reflects actual test performance. For example, if the critical level is in the lower end of the dynamic range, a bias estimate should also be obtained from the same concentration range, or the regression equation from a comparison of methods study can be used to estimate bias at the critical level. For tests with multiple critical levels, it may be desirable to make Sigma metric estimates at each level.

The relationship of imprecision and inaccuracy to Sigma metrics can be graphically depicted (Figure 2). Given a normal distribution of test results, as well as a known standard deviation (the imprecision) and a known bias, the acceptable performance range can easily be calculated and, conversely, the concentration ranges in which results are unacceptable can also be defined (i.e., the concentration ranges above and below the tolerance limits that define TEa).

The relationship between imprecision and inaccuracy to Sigma metrics can be summarized mathematically by the following equation:

Sigma-metric = (TEa – bias $_{observed}$)/CV$_{observed}$

A simple example with the Sigma-metric equation reveals that the "state of the art" in healthcare is not Six Sigma. For cholesterol, CLIA defines an allowable total error of 10%. That is, a cholesterol test result must be within 10% of its true value. The National Cholesterol Education Program (NCEP) established separate goals for imprecision and inaccuracy of 3% each. A method that performs with 3% CV and 3% bias is considered acceptable by the NCEP. The Sigma metric calculations tell another story:

$$(10 - 3)/3 = 2.33 \text{ Sigma}$$

This is stark proof that laboratories currently operate in an environment in which world class performance is not the goal.

Returning once again to the arrow and target model, the Sigma approach gives us a target, the quality requirement gives us the size of that target, and the performance data of the method give's us the arrow, which should land as close to the bull's-eye (Six Sigma) as possible.

Even when data on method bias is missing, a modified Sigma metric can be calculated. The resulting metric documents the capability of the method to achieve world class performance under ideal conditions when no bias is present. Since the real laboratory always operates with some amount of bias, the performance observed will always be lower than the Sigma capability. The benefit of such an assessment is that it allows the laboratory to estimate how much "room for error" is left after accounting for imprecision. For some instruments, even a Sigma capability metric will allow a laboratory to make judgments on the suitability of methods.

Sigma metric capability = (TE$_a$)/CV$_{observed}$

The performance of methods can be graphically illustrated using a Method Evaluation Decision chart [6] (MEDx, Figure 4) with Six Sigma metric lines imposed upon them. The Method Decision chart displays inaccuracy on the y-axis, imprecision on the x-axis. Typically the chart is drawn for each specific quality requirement (i.e., a 10% quality requirement would use a Method Decision chart drawn for 10%), but multiple methods with different quality requirements can be displayed on a Normalized Operating Specifications (OPSpecs) chart. In a Normalized Method Decision chart, the axes are each set to 100% and the x and y values are determined for the test by calculating its percentage of the quality requirement (Figure 5).

For example, if a test had a quality requirement of 10% and a CV of 1% and a bias of 2%, the coordinates on a Normalized Method Decision chart would be (10, 20) (Figure 6).
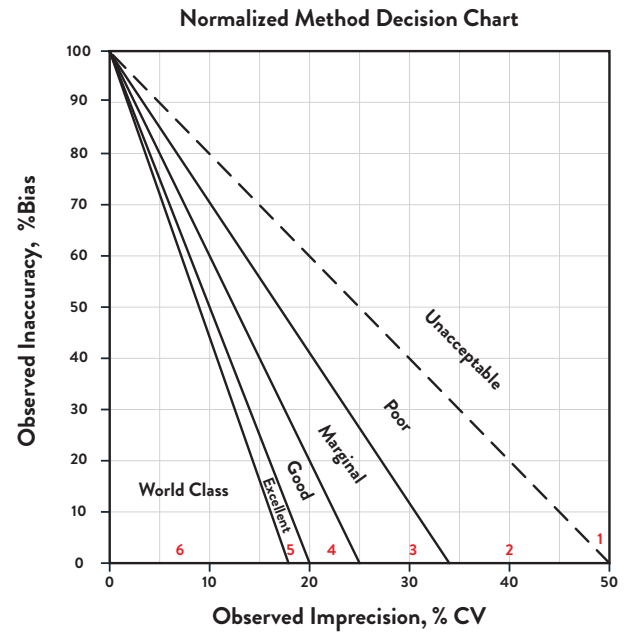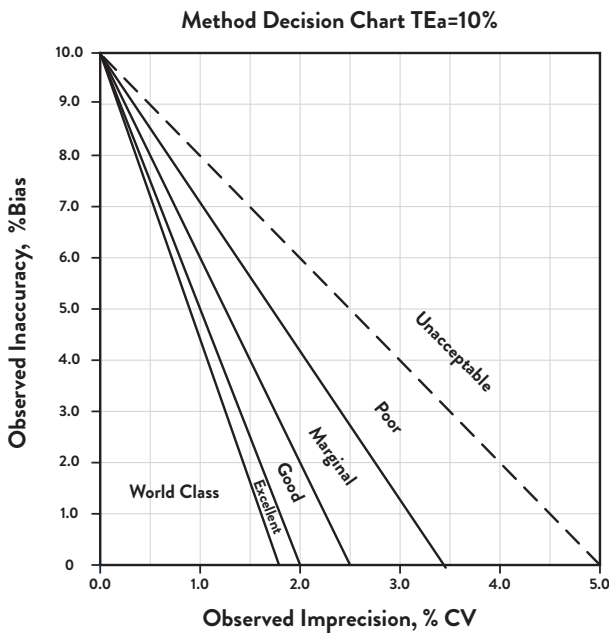
Figure 5: Normalized Method Decision chart

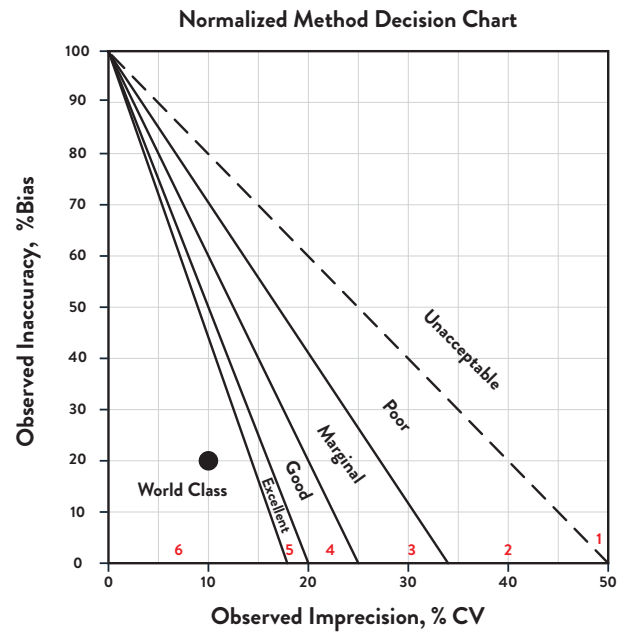**Figure 4**: Method Decision chart for an allowable total error of 10%

Figure 6: Normalized Method Decision chart with sample data point

While Normalized Method Decision charts with Six Sigma limits incorporate many complex features and calculations into a single display, the result of the chart still fits within the arrow and target model. The chart can be visualized as the upper right quadrant of the target. The area around the origin (0,0) of the chart (and below all of the lines) is the bull's eye. The Sigma lines drawn on the chart are similar to the rings of the target, with 3 Sigma representng the edge of the target (anything below 3 Sigma is considered off the target, i.e., unacceptable). The x- and y- coordinates of a plotted test represent the performance of the test and the spot where the arrow "landed."

Benefits of Sigma Assessment

Given the simple parameters of the Sigma-metric equation, laboratories can easily determine the current performance of all their current methods. The data acquired during the standard method validation protocols of a new instrument can also be used to determine performance metrics. In addition, formal method validation studies are typically conducted on all new methods and are often presented as posters at scientific conferences, as more scholarly reports published in scientific journals, or simply as data provided by the manufacturer upon request. With that data, laboratories can calculate Sigma metrics, compare them to the Sigma metrics of competing instruments, and use this tool as part of their decision-making process. This application, to objectively assess and compare instrument performance before purchasing a new instrument, is of considerable value. It gives the laboratory the power to predict which methods will perform to their clinical needs and which will not.

One Step Further: From Sigma Metric Assessment to Quality Control Design

Sigma metric analysis is not confined to the role of assessment and method validation. Sigma metrics can also be used to refine and streamline the operational routines of a method. Combining Sigma metrics with QC design tools, such as the Operating Specifications chart (OPSpecs), allows the laboratory to customize and optimize the QC procedures conducted by the laboratory. A rational QC Design can eliminate, much if not all, of the wasteful 2s QC practices, replacing them instead with appropriate control limits and numbers of control measurements.

An OPSpecs chart provides a graphic description of the imprecision and inaccuracy that are allowable and the

control rules and number of control measurements that are necessary for a QC procedure to achieve an appropriate level of analytical quality assurance for a defined quality requirement.[7] The diagonal lines in this chart represent the error detection performance of actual QC procedures (control rules and numbers of measurements). These lines are arranged from top to bottom according to their error detection capability; the highest line provides the highest error detection (thus, there is more "room" beneath that line for the method to hit). Other details about the QC procedure are noted in the key on the right side of the chart, such as the false rejection (Pfr) and number of control measurements (N) and the number of runs (R). The imprecision and inaccuracy of a method are used as the x-coordinate and y-coordinate, respectively. If this "operating point" lies below one of the lines of the OPSpecs chart, that indicates that the QC procedure represented by that line will provide the appropriate performance (Figure 7).

Reprising the arrow and target model one final time, the OPSpecs chart can be viewed in the same way as the Sigma metric analysis and the Method Decision chart. The OPSpecs chart is like the upper right quadrant of the target, with the origin as the bull's-eye. Method performance (the arrow) should be as close as possible to the bull's-eye. This time, however, the different rings on the target represent different QC procedures for use in the laboratory. The closer to the bull's-eye, the more QC procedures available for quality management.

OPSpecs charts, as with Method Decision charts, are typically generated for specific quality requirements. But OPSpecs charts can also be normalized so that multiple tests with different quality requirements can be displayed on the same chart (Figure 8).

Conclusion

Sigma metric analysis, Method Decision charts, and OPSpecs charts provide easy tools for laboratories to determine the performance of their current methods and QC design, and to compare competing instruments on the market. Both quantitative calculations and visual assessment can be made with this approach. These techniques give the laboratory a practical way to select the right method, and then select the right QC for that method. The result is an optimized testing process that fulfills the quality required for appropriate test interpretation.
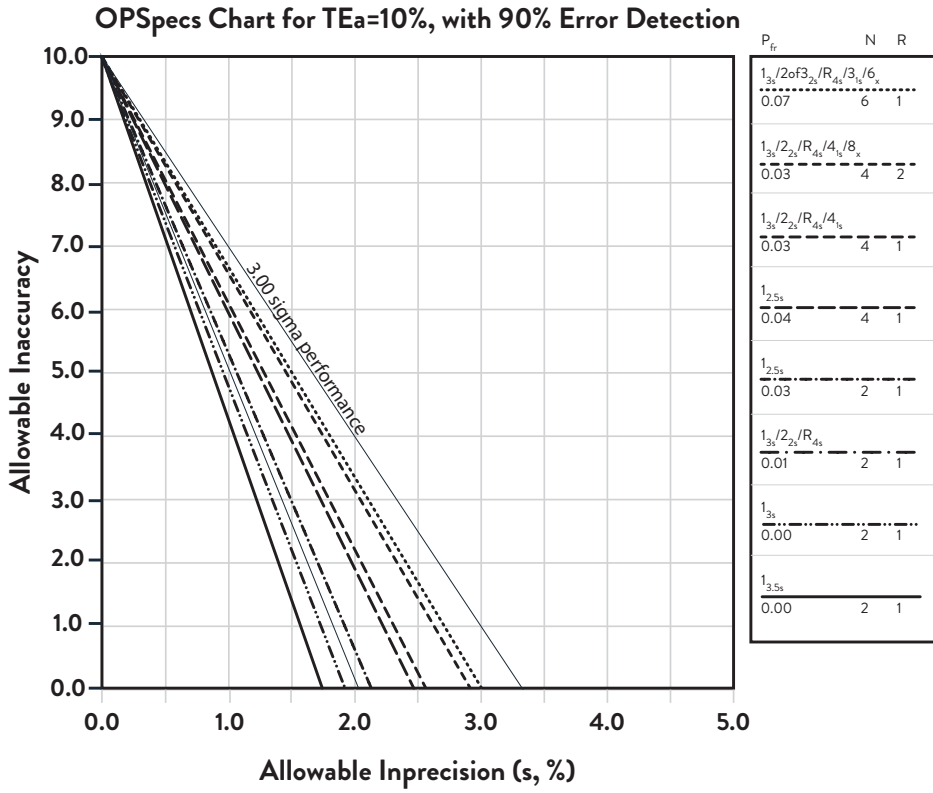
OPSpecs Chart for TEa=10%, with 90% Error Detection

Figure 7: OPSpecs chart for an allowable total error of 10%

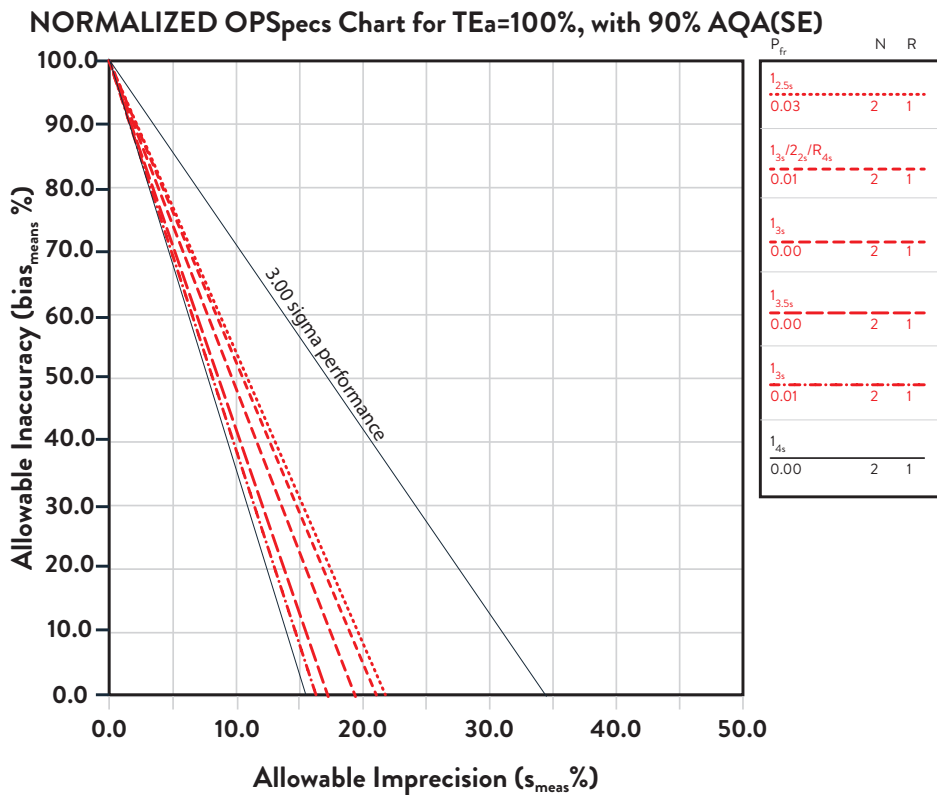## NORMALIZED OPSpecs Chart for TEa=100%, with 90% AQA(SE)

Figure 8: Sample Normalized OPSpecs chart

References

1. Westgard JO, *Six Sigma QC Design and Control*, 2nd Edition, Madison WI: Westgard QC, 2006.

2. Institute of Medicine. *To Err is Human: Building a Safer Health System*. Washington DC: National Academy Press, 2001.

3. Nevalainen D, Berte L, Kraft C, Leigh E, Morgan T. Evaluating laboratory performance on quality indicators with the Six Sigma scale. Arch Pathol Lab Med 2000;124:516-519. 4 Westgard JO, *Basic Planning for Quality*, Madison, Wisconsin: Westgard QC, 2000

4. Westgard JO, *Basic Planning for Quality*, Madison, Wisconsin: Westgard QC, 2000

5. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, Minchinela J, Perich C, Simon M. Desirable specifications for Total Error, Imprecision and Bias derived from Biologic Variation, http://www.westgard.com/biodatabase1.htm Accessed December 12, 2008.

6. Westgard JO. A Method Evaluation Decision Chart (MEDX chart) for judging method performance, *Clinical Laboratory Science*, 1995, vol. 8, no. 5, pp. 277-283

7. Westgard JO, Charts of operational process specifications ("OPSpecs charts") for assessing the precision, accuracy, and quality control needed to satisfy proficiency testing performance criteria *Clinical Chemistry* 38: 1226-1233, 1992

CORELABORATORY.ABBOTT

Abbott