Designation: D6299 – 21

An American National Standard

Standard Practice for Applying Statistical Quality Assurance and Control Charting Techniques to Evaluate Analytical Measurement System Performance[1]

This standard is issued under the fixed designation D6299; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope*

1.1 This practice covers information for the design and operation of a program to monitor and control ongoing stability and precision and bias performance of selected analytical measurement systems using a collection of generally accepted statistical quality control (SQC) procedures and tools.

NOTE 1—A complete list of criteria for selecting measurement systems to which this practice should be applied and for determining the frequency at which it should be applied is beyond the scope of this practice. However, some factors to be considered include (*1*) frequency of use of the analytical measurement system, (*2*) criticality of the parameter being measured, (*3*) system stability and precision performance based on historical data, (*4*) business economics, and (*5*) regulatory, contractual, or test method requirements.

1.2 This practice is applicable to stable analytical measurement systems that produce results on a continuous numerical scale.

1.3 This practice is applicable to laboratory test methods.

1.4 This practice is applicable to validated process stream analyzers.

1.5 This practice is applicable to monitoring the differences between two analytical measurement systems that purport to measure the same property provided that both systems have been assessed in accordance with the statistical methodology in Practice D6708 and the appropriate bias applied.

NOTE 2—For validation of univariate process stream analyzers, see also Practice D3764.

NOTE 3—One or both of the analytical systems in 1.5 may be laboratory test methods or validated process stream analyzers.

1.6 This practice assumes that the normal (Gaussian) model is adequate for the description and prediction of measurement system behavior when it is in a state of statistical control.

NOTE 4—For non-Gaussian processes, transformations of test results may permit proper application of these tools. Consult a statistician for further guidance and information.

1.7 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D3764 Practice for Validation of the Performance of Process Stream Analyzer Systems

D4175 Terminology Relating to Petroleum Products, Liquid Fuels, and Lubricants

D5191 Test Method for Vapor Pressure of Petroleum Products and Liquid Fuels (Mini Method)

D6300 Practice for Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products, Liquid Fuels, and Lubricants

D6617 Practice for Laboratory Bias Detection Using Single Test Result from Standard Material

D6708 Practice for Statistical Assessment and Improvement of Expected Agreement Between Two Test Methods that Purport to Measure the Same Property of a Material

D6792 Practice for Quality Management Systems in Petroleum Products, Liquid Fuels, and Lubricants Testing Laboratories

D7372 Guide for Analysis and Interpretation of Proficiency Test Program Results

D7915 Practice for Application of Generalized Extreme Studentized Deviate (GESD) Technique to Simultaneously Identify Multiple Outliers in a Data Set

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

E178 Practice for Dealing With Outlying Observations

[1] This practice is under the jurisdiction of ASTM Committee D02 on Petroleum Products, Liquid Fuels, and Lubricants and is the direct responsibility of Subcommittee D02.94 on Coordinating Subcommittee on Quality Assurance and Statistics.

Current edition approved May 1, 2021. Published June 2021. Originally approved in 1998. Last previous edition approved in 2020 as D6299 – 20a. DOI: 10.1520/D6299-21.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

***A Summary of Changes section appears at the end of this standard**

E456 Terminology Relating to Quality and Statistics
E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

3. Terminology

3.1 *Definitions:*

3.1.1 More extensive lists of terms related to quality and statistics are found in Terminology D4175, Practice D6300, and Terminology E456.

3.1.2 *repeatability conditions, n*—conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. **D6300**

3.1.3 *reproducibility (R), n*—a quantitative expression for the random error associated with the difference between two independent results obtained under reproducibility conditions that would be exceeded with an approximate probability of 5 % (one case in 20 in the long run) in the normal and correct operation of the test method. **D6300**

3.1.4 *reproducibility conditions, n*—conditions where independent test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

3.1.4.1 *Discussion*—Different laboratory by necessity means a different operator, different equipment, and different location and under different supervisory control. **D6300**

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 More extensive lists of terms related to quality and statistics are found in Terminology D4175, Practice D6300, and Terminology E456.

3.2.2 *accepted reference value, n*—a value that serves as an agreed-upon reference for comparison and that is derived as (*1*) a theoretical or established value, based on scientific principles, (*2*) an assigned value, based on experimental work of some national or international organization, such as the U.S. National Institute of Standards and Technology (NIST), or (*3*) a consensus value, based on collaborative experimental work under the auspices of a scientific or engineering group.

3.2.3 *accuracy, n*—the closeness of agreement between an observed value and an accepted reference value.

3.2.4 *analytical measurement system, n*—a collection of one or more components or subsystems, such as samplers, test equipment, instrumentation, display devices, data handlers, printouts or output transmitters, that is used to determine a quantitative value of a specific property for an unknown sample in accordance with a test method.

3.2.4.1 *Discussion*—A standard test method (for example, ASTM, ISO) may be an example of an *analytical measurement system*.

3.2.4.2 *Discussion*—The control chart methodology and work processes described in this practice are intended to be applied independently to the final results produced from each individual measurement system, or, differences between results from two individual measurement systems for the same test sample. They are not intended to be applied to combined final results from multiple individual analytical systems or different instruments executing the same test method.

3.2.5 *assignable cause, n*—a factor that contributes to variation and that is feasible to detect and identify.

3.2.6 *bias, n*—a systematic error that contributes to the difference between a population mean of the measurements or test results and an accepted reference or true value.

3.2.7 *blind submission, n*—submission of a check standard or quality control (QC) sample for analysis without revealing the expected value to the person performing the analysis.

3.2.8 *check standard, n—in QC testing*, a material having an accepted reference value used to determine the accuracy of a measurement system.

3.2.8.1 *Discussion*—A check standard is preferably a material that is either a certified reference material with traceability to a nationally recognized body or a material that has an accepted reference value established through interlaboratory testing. For some measurement systems, a pure, single component material having known value or a simple gravimetric or volumetric mixture of pure components having calculable value may serve as a check standard. Users should be aware that for measurement systems that show matrix dependencies, accuracy determined from pure compounds or simple mixtures may not be representative of that achieved on actual samples.

3.2.9 *common (chance, random) cause, n*—for quality assurance programs, one of generally numerous factors, individually of relatively small importance, that contributes to variation, and that is not feasible to detect and identify.

3.2.10 *control limits, n*—limits on a control chart that are used as criteria for signaling the need for action or for judging whether a set of data does or does not indicate a state of statistical control.

3.2.11 *double blind submission, n*—submission of a check standard or QC sample for analysis without revealing the check standard or QC sample status and expected value to the person performing the analysis.

3.2.12 *in-statistical-control, adj*—a process, analytical measurement system, or function that exhibits variations that can only be attributable to common cause.

3.2.13 *lot, n*—a definite quantity of a product or material accumulated under conditions that are considered uniform for sampling purposes.

3.2.14 *out-of-statistical-control, adj*—a process, analytical measurement system, or function that exhibits variations in addition to those that can be attributable to common cause and the magnitude of these additional variations exceed specified limits.

3.2.14.1 *Discussion*—For clarification, a transition from an in-statistical-control system to an out-of-statistical-control system does not automatically imply that there is a change in the fit for use status of the system in terms of meeting the requirements for the intended application.

3.2.15 *precision, n*—the closeness of agreement between test results obtained under prescribed conditions.

3.2.16 *proficiency testing, n*—determination of a laboratory's testing capability by participation in an interlaboratory crosscheck program.

3.2.16.1 *Discussion*—ASTM Committee D02 conducts proficiency testing among hundreds of laboratories, using a wide variety of petroleum products and lubricants.

3.2.17 *quality control (QC) sample, n*—for use in quality assurance programs to determine and monitor the precision and stability of a measurement system, a stable and homogeneous material having physical or chemical properties, or both, similar to those of typical samples tested by the analytical measurement system. The material is properly stored to ensure sample integrity, and is available in sufficient quantity for repeated, long term testing.

3.2.18 *system expected value (SEV), n*—for a QC sample this is an estimate of the theoretical limiting value towards which the average of results collected from a single in-statistical-control measurement system under site precision conditions tends as the number of results approaches infinity.

3.2.18.1 *Discussion*—The SEV is associated with a single measurement system; for control charts that are plotted in actual measured units, the SEV is required, since it is used as a reference value from which upper and lower control limits for the control chart specific to a batch of QC material are constructed.

3.2.19 *site precision (R'), n*—the value which the absolute difference between two individual test results obtained under site precision conditions is expected to exceed about 5 % of the time (one case in 20 in the long run) in the normal and correct operation of the test method.

3.2.19.1 *Discussion*—It is defined as 2.77 times $\sigma_{R'}$, the standard deviation of results obtained under site precision conditions.

3.2.20 *site precision conditions, n*—conditions under which test results are obtained by one or more operators in a single site location practicing the same test method on a single measurement system which may comprise multiple instruments, using test specimens taken at random from the same sample of material, over an extended period of time spanning at least a 15 day interval.

3.2.20.1 *Discussion*—Site precision conditions should include all sources of variation that are typically encountered during normal, long term operation of the measurement system. Thus, all operators who are involved in the routine use of the measurement system should contribute results to the site precision determination. In situations of high usage of a test method where multiple QC results are obtained within a 24 h period, then only results separated by at least 4 h to 8 h, depending on the absence of auto-correlation in the data, the nature of the test method/instrument, site requirements, or regulations, should be used in site precision calculations to reflect the longer term variation in the system.

3.2.21 *site precision standard deviation, n*—the standard deviation of results obtained under site precision conditions for an individual measurement system and materials that are similar in composition and property level to the QC samples used to establish the standard deviation.

3.2.22 *validation audit sample, n*—a QC sample or check standard used to verify precision and bias estimated from routine quality assurance testing.

3.3 *Symbols:*

3.3.1 *ARV*—accepted reference value.

3.3.2 *EWMA*—exponentially weighted moving average.

3.3.3 *I*—individual observation (as in *I*-chart).

3.3.4 *MR*—moving range.

3.3.5 $\overline{MR}$—average of moving range.

3.3.6 *QC*—quality control.

3.3.7 *R'*—site precision.

3.3.8 *SEV*—site expected value.

3.3.9 $\sigma_{R'}$—site precision standard deviation.

3.3.10 *VA*—validation audit.

3.3.11 χ^2—chi squared.

3.3.12 λ—lambda.

4. Summary of Practice

4.1 QC samples and check standards are regularly analyzed by the measurement system. Control charts and other statistical techniques are presented to screen, plot, and interpret test results in accordance with industry-accepted practices to ascertain the in-statistical-control status of the measurement system.

4.2 Statistical estimates of the measurement system precision and bias are calculated and periodically updated using accrued data.

4.3 In addition, as part of a separate validation audit procedure, QC samples and check standards may be submitted blind or double-blind and randomly to the measurement system for routine testing to verify that the calculated precision and bias are representative of routine measurement system performance when there is no prior knowledge of the expected value or sample status.

5. Significance and Use

5.1 This practice may be used to continuously demonstrate the proficiency of analytical measurement systems that are used for establishing and ensuring the quality of petroleum and petroleum products.

5.2 Data accrued, using the techniques included in this practice, provide the ability to monitor analytical measurement system precision and bias.

5.3 These data are useful for updating test methods as well as for indicating areas of potential measurement system improvement.

6. Reference Materials

6.1 QC samples are used to establish and monitor the precision of the analytical measurement system.

6.1.1 Select a stable and homogeneous material having physical or chemical properties, or both, similar to those of typical samples tested by the analytical measurement system.

NOTE 5—When the QC sample is to be utilized for monitoring a process stream analyzer performance, it is often helpful to supplement the process analyzer system with a subsystem to automate the extraction, mixing, storage, and delivery functions associated with the QC sample.

6.1.2 Estimate the quantity of the material needed for each specific lot of QC sample to (*1*) accommodate the number of analytical measurement systems for which it is to be used (laboratory test apparatuses as well as process stream analyzer systems) and (*2*) provide determination of QC statistics for a useful and desirable period of time.

6.1.3 Collect the material into a single container and isolate it.

6.1.4 Thoroughly mix the material to ensure homogeneity.

6.1.5 Conduct any testing necessary to ensure that the QC sample meets the characteristics for its intended use.

6.1.6 Package or store QC samples, or both, as appropriate for the specific analytical measurement system to ensure that all analyses of samples from a given lot are performed on essentially identical material. If necessary, split the bulk material collected in 6.1.3 into separate and smaller containers to help ensure integrity over time. (**Warning—**Treat the material appropriately to ensure its stability, integrity, and homogeneity over the time period for which it is to stored and used. For samples that are volatile, such as gasoline, storage in one large container that is repeatedly opened and closed may result in loss of light ends. This problem can be avoided by chilling and splitting the bulk sample into smaller containers, each with a quantity sufficient to conduct the analysis. Similarly, samples prone to oxidation may benefit from splitting the bulk sample into smaller containers that can be blanketed with an inert gas prior to being sealed and leaving them sealed until the sample is needed.)

6.2 Check standards are used to estimate the accuracy of the analytical measurement system.

6.2.1 A check standard may be a commercial standard reference material when such material is available in appropriate quantity, quality and composition.

NOTE 6—Commercial reference material of appropriate composition may not be available for all measurement systems.

6.2.2 Alternatively, a check standard may be prepared from a material that is analyzed under reproducibility conditions by multiple measurement systems. The accepted reference value (ARV) for this check standard shall be the average after statistical examination and outlier treatment has been applied.[3]

6.2.2.1 Exchange samples circulated as part of an interlaboratory exchange program, or round robin, may be used as check standards. For the average computed from an exchange sample to be usable as the Accepted Reference Value (ARV) of a check standard, the standard deviation computed from at least 16 non-rejected normally distributed results (single submission per participant) shall not be statistically greater than the reproducibility standard deviation for the test method. An *F*-test should be applied to test acceptability.

NOTE 7—The uncertainty in the ARV is inversely proportional to the square root of the number of values in the average. For example, use of 16 non-outlier results in calculating the ARV reduces the uncertainty of the ARV by a factor of 4 relative to the single result precision. The bias tests described in this practice assume that the uncertainty in the ARV is negligible relative to the precision of the measurement system being evaluated. If less than 16 values are used in calculating the average, this assumption may not be valid.

NOTE 8—Examples of exchanges that may be acceptable are ASTM D02.92 Proficiency Test Program; ASTM D02.01 N.E.G.; ASTM D02.01.A Regional Exchanges; International Quality Assurance Exchange Program, administered by Innotech ALBERTA.

6.2.3 For some measurement systems, single, pure component materials with known value, or simple gravimetric or volumetric mixtures of pure components having calculable value may serve as a check standard. For example, pure solvents, such as 2,2-dimethylbutane, are used as check standards for the measurement of Reid vapor pressure by Test Method D5191. Users should be aware that for measurement systems that show matrix dependencies, accuracy determined from pure compounds or simple mixtures may not be representative of that achieved on actual samples.

6.3 Validation audit (VA) samples are QC samples and check standards, which may, at the option of the users, be submitted to the measurement system in a blind, or double blind, and random fashion to verify precision and bias estimated from routine quality assurance testing.

7. Quality Assurance (QA) Program for Individual Measurement Systems

7.1 *Overview*—A QA program (**1**)[4] may consist of five primary activities: (*1*) monitoring stability and precision through QC sample testing, (*2*) monitoring accuracy, (*3*) periodic evaluation of system performance in terms of precision or bias, or both, (*4*) proficiency testing through participation in interlaboratory exchange programs where such programs are available, and (*5*) a periodic and independent system validation using VA samples may be conducted to provide additional assurance of the system precision and bias metrics established from the primary testing activities. At minimum, the QA program must include at least item one and item two, subject to check standard availability (see 7.1.1).

7.1.1 For some measurement systems, suitable check standard materials may not exist, and there may be no reasonably available exchange programs to generate them. For such systems, there is no means of verifying the accuracy of the system, and the QA program will only involve monitoring stability and precision through QC sample testing.

NOTE 9—For guidance on the establishment and maintenance of the essentials of a quality system, see Practice D6792.

NOTE 10—For guidance on the analysis and interpretation of proficiency test (PT) program results, see Guide D7372.

7.2 *Monitoring System Stability and Precision Through QC Sample Testing*—QC test specimen samples from a specific lot are introduced and tested in the analytical measurement system on a regular basis to establish system performance history in terms of both stability and precision.

[3] For guidance in statistical and outlier treatment of data, refer to Practices D6300, D7915, E178, and E691.

[4] The boldface numbers in parentheses refer to the list of references at the end of this standard.