Designation: D6617 – 08

An American National Standard

Standard Practice for Laboratory Bias Detection Using Single Test Result from Standard Material[1]

This standard is issued under the fixed designation D6617; 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.

INTRODUCTION

Due to the inherent imprecision in all test methods, a laboratory cannot expect to obtain the numerically exact accepted reference value (ARV) of a check standard (CS) material every time one is tested. Results that are reasonably close to the ARV should provide assurance that the laboratory is performing the test method either without bias, or with a bias that is of no practical concern, hence requiring no intervention. Results differing from the ARV by more than a certain amount, however, should lead the laboratory to take corrective action.

1. Scope*

1.1 This practice covers a methodology for establishing an acceptable tolerance zone for the difference between the result obtained from a single implementation of a test method on a CS and its ARV, based on user-specified Type I error, the user-established test method precision, the standard error of the ARV, and a presumed hypothesis that the laboratory is performing the test method without bias.

NOTE 1—Throughout this practice, the term user refers to the user of this practice; and the term laboratory (see 1.1) refers to the organization or entity that is performing the test method.

1.2 For the tolerance zone established in 1.1, a methodology is presented to estimate the probability that the single test result will fall outside the zone, in the event that there is a bias (positive or negative) of a user-specified magnitude that is deemed to be of practical concern (that is, the presumed hypothesis is not true).

1.3 This practice is intended for ASTM Committee D02 test methods that produce results on a continuous numerical scale.

1.4 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 2—While this practice does not cover scenarios in which multiple results are obtained on the same CS under site precision or repeatability conditions, the statistical concepts presented are applicable. Users wishing to apply these concepts for the scenarios described are advised to consult a statistician and to reference the CS methodology described in Practice D6299.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D2699 Test Method for Research Octane Number of Spark-Ignition Engine Fuel

D6299 Practice for Applying Statistical Quality Assurance and Control Charting Techniques to Evaluate Analytical Measurement System Performance

E178 Practice for Dealing With Outlying Observations

3. Terminology

3.1 Definitions for accepted reference value (ARV), accuracy, bias, check standard (CS), in statistical control, site precision, site precision standard deviation (σ_{SITE}), site precision conditions, repeatability conditions, and reproducibility conditions can be found in Practice D6299.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *acceptable tolerance zone, n*—a numerical zone bounded inclusively by zero $\pm k\epsilon$ (k is a value based on a user-specified Type I error; ϵ is defined in 3.2.7) such that if the difference between the result obtained from a single implementation of a test method for a CS and its ARV falls inside this zone, the presumed hypothesis that the laboratory or testing organization is performing the test method without bias is accepted, and the difference is attributed to normal random

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

Current edition approved Dec. 1, 2008. Published January 2009. Originally approved in 2000. Last previous edition approved in 2005 as D6617–05. DOI: 10.1520/D6617-08.

[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 variation of the test method. Conversely, if the difference falls outside this zone, the presumed hypothesis is rejected.

3.2.2 *consensus check standard (CCS), n*— a special type of CS in which the ARV is assigned as the arithmetic average of at least 16 non-outlying (see Practice E178 or equivalent) test results obtained under reproducibility conditions, and the results pass the Anderson-Darling normality test in Practice D6299, or other statistical normality test at the 95 % confidence level.

3.2.2.1 *Discussion*—These may be production materials with unspecified composition, but are compositionally representative of material routinely tested by the test method, or materials with specified compositions that are reproducible, but may not be representative of routinely tested materials.

3.2.3 *delta* (Δ), *n*—a signless quantity, to be specified by the user as the minimum magnitude of bias (either positive or negative) that is of practical concern.

3.2.4 *power of bias detection, n*—in applying the methodology of this practice, this refers to the long run probability of being able to correctly detect a bias of a magnitude of at least Δ; given the acceptance tolerance zone set under the presumed hypothesis, and is defined as (1 – Type II error), for a user-specified Δ.

3.2.4.1 *Discussion*—The quantity (1 – Type II error), commonly known as the power of the test in classical statistical hypothesis testing, refers to the probability of correctly rejecting the null hypothesis, given that the alternate hypothesis is true. In applying this SP, the power refers to the probability of detecting a positive or negative bias of at least Δ.

3.2.5 *standardized delta* (Δ_S), *n*—Δ, expressed in units of total uncertainty (ε) per the equation:

$$(\Delta_S) = \Delta/\epsilon \quad (1)$$

3.2.6 *standard error of ARV* (SE_{ARV}), *n*—a statistic quantifying the uncertainty associated with the ARV in which the latter is used as an estimate for the true value of the property of interest. For a CCS, this is defined as:

$$\sigma_{CCS}/\sqrt{N} \quad (2)$$

where:

N = total number of non-outlying results used to establish the ARV, collected under reproducibility conditions, and

σ_{CCS} = the standard deviation of all the non-outlying results.

3.2.6.1 *Discussion*—Assuming a normal model, a 95 % confidence interval that would contain the true value of the property of interest can be constructed as follows:

$$ARV - 1.96\, SE_{ARV} \text{ to } ARV + 1.96\, SE_{ARV} \quad (3)$$

3.2.7 *total uncertainty* (ε), *n*—combined quantity of test method σ_{SITE} and SE_{ARV} as follows:

$$\epsilon = \sqrt{\sigma^2_{SITE} + SE^2_{ARV}} \quad (4)$$

3.2.8 *type I error, n*—in applying the methodology of this practice, this refers to the theoretical long run probability of rejecting the presumed hypothesis that the test method is performed without bias when in fact the hypothesis is true, hence, committing an error in decision.

3.2.8.1 *Discussion*—Type I error, commonly known as alpha (α) error in classical statistical hypothesis testing, refers to the probability of incorrectly rejecting a presumed, or null hypothesis based on statistics generated from relevant data. In applying this practice, the null hypothesis is stated as: The test method is being performed without bias; or it can be equivalently stated as: H_0: bias = 0.

3.2.9 *type II error, n*—in applying the methodology of this practice, this refers to the long run probability of accepting (that is, not rejecting) the presumed hypothesis that the method is performed without bias, when in fact the presumed hypothesis is not true, and the test method is biased by a magnitude of at least Δ, hence, committing an error in decision.

3.2.9.1 *Discussion*—Type II error, commonly known as beta (β) error in classical statistical hypothesis testing, refers to the probability of failure to reject the null hypothesis when it is not true, based on statistics generated from relevant data. To quantify Type II error, the user is required to declare a specific alternate hypothesis that is believed to be true. In applying this practice, the alternate hypothesis will take the form: "The test method is biased by at least Δ", where Δ is *a priori* decided by the user as the minimum amount of bias in either direction (positive or negative) that is of practical concern. The alternate hypothesis can be equivalently stated as: H_1: $|\text{bias}| \geq \Delta$.

4. Significance and Use

4.1 Laboratories performing petroleum test methods can use this practice to set an acceptable tolerance zone for infrequent testing of CS or CCS material, based on ϵ, and a desired Type I error, for the purpose of ascertaining if the test method is being performed without bias.

4.2 This practice can be used to estimate the power of correctly detecting bias of different magnitudes, using the acceptable tolerance zone set in 4.1, and hence, gain insight into the limitation of the true bias detection capability associated with this acceptable tolerance zone. With this insight, trade-offs can be made between desired Type I error versus desired bias detection capability to suit specific business needs.

4.3 The CS testing activities described in this practice are intended to augment and not replace the regular statistical monitoring of test method performance as described in Practice D6299.

5. General Requirement

5.1 Application of the methodology in this practice requires the following:

5.1.1 The standard material has an ARV and associated standard error (SE_{ARV}).

NOTE 3—For a given power of detection, the magnitude of the associated bias detectable is directly proportional to $\epsilon = \sqrt{SE^2_{ARV} + \sigma^2_{SITE}}$. Therefore, efforts should be made to keep the ratio (SE_{ARV} / σ_{SITE}) to as low a value as practical. A ratio of 0.5 or less is considered useful.

5.1.2 The user has a σ_{SITE} for the test method that is reasonably suited for the standard material.

NOTE 4—It is recognized that there will be situations in which the CS may not be compositionally similar to or have property level similar to, or