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 D 6617; 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 D 6299.

2. Referenced Documents

2.1 *ASTM Standards:*

D 2699 Test Method for Research Octane Number of Spark-Ignition Engine Fuel[2]

D 6299 Practice for Applying Statistical Quality Assurance Techniques to Evaluate Analytical Measurement System Performance[3]

E 178 Practice for Dealing with Outlying Observations[4]

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 D 6299.

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 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 E 178 or equivalent) test

[1] This practice is under the jurisdiction of ASTM Committee D02 on Petroleum Products and Lubricants and is the direct responsibility of Subcommittee CS94 on Quality Assurance and Statistics.

Current edition approved Dec. 10, 2000. Published February 2001.

[2] *Annual Book of ASTM Standards*, Vol 05.01.

[3] *Annual Book of ASTM Standards*, Vol 05.03.

[4] *Annual Book of ASTM Standards*, Vol 14.02.

results obtained under reproducibility conditions, and the results pass the Anderson-Darling normality test in Practice D 6299, 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:

$$(\Delta_S) = \Delta / \epsilon \tag{1}$$

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:

$$\sigma_{CCS} / \sqrt{N} \tag{2}$$

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:

$$ARV - 1.96\ SE_{ARV} \text{ to } ARV + 1.96\ SE_{ARV} \tag{3}$$

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

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

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 D 6299.

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 both, the materials regularly tested. For those situations, the site precision standard deviation (σ_{SITE}) estimated using regularly tested material at a property level closest to the check standard should be used.

5.1.3 The user should pre-specify the required Type I error and the minimum magnitude of bias that is of practical concern (Δ).

5.1.4 The test method is in statistical control.

NOTE 5—Within the context of this practice, a test method can be in statistical control (that is, mean is stable, under common cause variations), but can be biased.

6. Procedure

6.1 Confirm the usefulness of the CS by assessing the ratio $[SE_{ARV} \ / \ \sigma_{SITE}]$.

NOTE 6—A ratio of less than or equal to 0.5 is considered useful.

6.2 Calculate $\epsilon \ = \ \sqrt{\sigma^2_{SITE} + SE^2_{ARV}}$

6.3 Specify the required Type I error rate.

NOTE 7—A suggested starting value is 0.05.

6.4 Specify required Δ.

NOTE 8—The magnitude of Δ is usually specified based on nonstatistical considerations such as business risks or operational issues, or both.

6.5 Calculate $\Delta_S \ = \ \Delta / \epsilon$.

6.6 See Table 1.

6.7 Look across the row with the Δ_S values and identify the column with a Δ_S value closest to the Δ_S calculated in 6.5.

6.8 Look down the column identified in 6.7 and locate the row with the value in Column A closest to the required Type I error. The value in the cell where the row and column intersect is the power of detection.

6.9 If the power of detection is not acceptable (typically it will be too low), iteratively change one or all of the following until all requirements are met.

6.9.1 Type I error.

6.9.2 Delta (Δ).

6.9.3 Power of bias detection.

NOTE 9—For a single implementation of the test method, the power of bias detection will depend on the magnitude of Δ specified, the total uncertainty ϵ, and the specified Type I error rate. Power of bias detection will increase at the expense of an increase in Type I error rate or increase in Δ.

6.10 Use the appropriate *k* value from Column B of Table 1 that met the specified Type I error and power of bias detection to calculate the boundaries of the acceptable tolerance zone.

6.11 Construct the acceptable tolerance zone: $0 \pm k\epsilon$.

6.12 When a single test result *X* for a CS is obtained, calculate the quantity $(X - ARV)$.

6.13 If $X - ARV$ falls inside the acceptable tolerance zone inclusively, accept the presumed hypothesis that the laboratory is performing the test method without bias.

TABLE 1 Type I Error and Associated Power of Bias Detection for Various Δ_s Values

A	B	Power of Detecting $\|bias\| = \Delta_s$[A]												
Required Type I Error Rate	k	(Δ_s) =0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	4
0.05	1.96	0.07	0.11	0.17	0.24	0.32	0.42	0.52	0.61	0.71	0.79	0.85	0.94	0.98
0.10	1.64	0.13	0.19	0.26	0.35	0.44	0.54	0.64	0.73	0.80	0.87	0.91	0.97	0.99
0.15	1.44	0.17	0.25	0.33	0.42	0.52	0.62	0.71	0.79	0.86	0.90	0.94	0.98	0.99
0.2	1.28	0.22	0.30	0.39	0.49	0.59	0.68	0.76	0.83	0.89	0.93	0.96	0.99	1.00
0.25	1.15	0.26	0.34	0.44	0.54	0.64	0.73	0.80	0.86	0.91	0.95	0.97	0.99	1.00
0.30	1.04	0.30	0.39	0.49	0.58	0.68	0.76	0.83	0.89	0.93	0.96	0.98	0.99	1.00
0.35	0.93	0.33	0.43	0.53	0.62	0.71	0.79	0.86	0.91	0.94	0.97	0.98	0.99	1.00
0.4	0.84	0.37	0.46	0.56	0.66	0.74	0.82	0.88	0.92	0.95	0.97	0.98	1.00	1.00
0.45	0.76	0.40	0.50	0.60	0.69	0.77	0.84	0.89	0.93	0.96	0.98	0.99	1.00	1.00
0.50	0.67	0.43	0.53	0.63	0.72	0.80	0.86	0.91	0.94	0.97	0.98	0.99	1.00	1.00
0.55	0.60	0.46	0.56	0.66	0.74	0.82	0.88	0.92	0.95	0.97	0.98	0.99	1.00	1.00
0.6	0.52	0.49	0.59	0.68	0.77	0.84	0.89	0.93	0.96	0.98	0.99	0.99	1.00	1.00
0.65	0.45	0.52	0.62	0.71	0.79	0.85	0.90	0.94	0.96	0.98	0.99	0.99	1.00	1.00
0.70	0.39	0.55	0.64	0.73	0.81	0.87	0.91	0.95	0.97	0.98	0.99	1.00	1.00	1.00
0.75	0.32	0.57	0.67	0.75	0.82	0.88	0.92	0.95	0.97	0.99	0.99	1.00	1.00	1.00
0.80	0.25	0.60	0.69	0.77	0.84	0.89	0.93	0.96	0.98	0.99	0.99	1.00	1.00	1.00
0.85	0.19	0.62	0.71	0.79	0.86	0.91	0.94	0.96	0.98	0.99	0.99	1.00	1.00	1.00
0.90	0.13	0.65	0.73	0.81	0.87	0.92	0.95	0.97	0.98	0.99	1.00	1.00	1.00	1.00

[A] Δ_s = Power of detecting bias.

6.14 If $X - ARV$ falls outside the acceptable tolerance zone on the positive side, reject the presumed hypothesis that the laboratory is performing the test method without bias, and conclude that there is evidence to suggest the laboratory is performing the test method with a positive bias of at least the magnitude Δ.

6.15 If $X - ARV$ falls outside the acceptable tolerance zone on the negative side, reject the presumed hypothesis that the laboratory is performing the test method without bias, and conclude that there is evidence to suggest the laboratory is performing the test method with a negative bias of at least the magnitude Δ.

7. Keywords

7.1 accepted reference value; bias; check standard; consensus; power of test; probability of bias detection; type I error; type II error

APPENDIX

(Nonmandatory Information)

X1. WORKED EXAMPLE

X1.1 The purpose of this appendix is to provide a worked example of the proper execution of the methodology described in this practice.

X1.2 A laboratory wishes to apply this practice to assess if it is performing Test Method D 2699 (Research Octane Number) without bias by a single test of a gasoline CCS with an ARV of 92.2.

X1.3 Details of the CCS are as follows:

X1.3.1 A total of 30 laboratories participated in an interlaboratory exchange program and each lab tested this gasoline once.

X1.3.2 The standard deviation from all the nonoutlying results is 0.25.

X1.3.3 The Anderson-Darling test for normality is not significant at the 95 % confidence level.

X1.4 Following definitions and procedures prescribed in this practice (shown in *italics* below):

X1.4.1 From 6.1: *Confirm the usefulness of the CS by assessing the ratio* $[SE_{ARV} / \sigma_{SITE}]$ *. A ratio of less than or equal to 0.5 is considered useful.*

X1.4.1.1 The σ_{SITE} for Test Method D 2699 is estimated at 0.1, using results from regular testing of an internal quality control production gasoline of 91.1 research octane.

X1.4.1.2 SE_{ARV} of the *CCS* $= 0.25 / \sqrt{30} = 0.046$.

X1.4.1.3 $[SE_{ARV} / \sigma_{SITE}] = 0.046 / 0.1 = 0.46 < 0.5$; conclude this CS is useful for assessing bias.

X1.4.2 From 6.2: *Calculate* ϵ:

$$\epsilon = \sqrt{0.1^2 + 0.046^2} = 0.11 \quad (X1.1)$$

X1.4.3 From 6.3: *Specify the required Type I error rate.* Following the suggestion of this SP, this is set at 0.05.

X1.4.4 From 6.4: *Specify required* Δ. Based on business requirement of the lab, Δ is set at $2\epsilon = 0.22$;

X1.4.5 From 6.5: *Calculate* $\Delta_S = \Delta / \epsilon$:

$$\Delta_S = 2 \quad (X1.2)$$

X1.4.6 From 6.6: *Go to Table 1.*

X1.4.7 From 6.7: *Look across the row with the* Δ_S *values and identify the column with a* Δ_S *value closest to the* Δ_S *calculated in 6.5.* Locate the column with Δ_S labeled 2.

X1.4.8 From 6.8: *Look down the column identified in 6.7 and locate the row with the value in column A closest to the required Type I error. The value in this cell is the power of detection.* The power of detection is 0.52.

X1.4.9 From 6.9: *If the power of detection is not acceptable, (typically it will be too low), iteratively change one or all of the following until all requirements are met:*

(1) Type I error,

(2) Δ*, and*

(3) power of bias detection.

X1.4.9.1 The laboratory decided that the power of detection is too low, and iteratively examined the values in the column identified in 6.7 versus the corresponding Type I errors in Column A. The laboratory ultimately accepted a 0.2 (20 %) Type I error rate in return for a higher power of bias detection (0.76, or 76 %) for their Δ.

X1.4.10 From 6.10: *Use the appropriate k value from Column B of Table 1 that met the specified Type I error and probability of bias detection to calculate the boundaries of the acceptable tolerance zone.* $k = 1.28$ (Column B).

X1.4.11 From 6.11: *Construct the acceptable tolerance zone:* $0 \pm k\epsilon$. The acceptable tolerance zone is $0 \pm 1.28 \times 0.11 = > -0.14$ *to* $+0.14$.

X1.4.12 From 6.12: *When a single test result X for a CS is obtained, calculate the quantity (X – ARV).* Suppose a result of 92.5 is obtained from a single implementation of this CCS, then $(X - 92.2) = 92.5 - 92.2 = 0.3$.

X1.4.13 From 6.13: *If (X – ARV) falls inside the acceptable tolerance zone inclusively, accept the presumed hypothesis that the laboratory is performing the test method with no bias that is of practical concern.* Since 0.3 falls outside the acceptable tolerance zone, the presumed hypothesis of no bias is rejected.

X1.4.14 From 6.14: *If (X – ARV) falls outside the acceptable tolerance zone on the positive side, reject the presumed hypothesis in 6.13, and conclude that there is evidence to suggest the laboratory is performing the test method with a* *positive bias by a magnitude of at least* Δ. It is concluded that the laboratory is performing this test method with a positive bias of at least 0.3.