



Designation: E2617 – 17

# Standard Practice for Validation of Empirically Derived Multivariate Calibrations<sup>1</sup>

This standard is issued under the fixed designation E2617; 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 ( $\epsilon$ ) indicates an editorial change since the last revision or reapproval.

## 1. Scope

1.1 This practice covers requirements for the validation of empirically derived calibrations (**Note 1**) such as calibrations derived by Multiple Linear Regression (MLR), Principal Component Regression (PCR), Partial Least Squares (PLS), Artificial Neural Networks (ANN), or any other empirical calibration technique whereby a relationship is postulated between a set of variables measured for a given sample under test and one or more physical, chemical, quality, or membership properties applicable to that sample.

**NOTE 1**—Empirically derived calibrations are sometimes referred to as “models” or “calibrations.” In the following text, for conciseness, the term “calibration” may be used instead of the full name of the procedure.

1.2 This practice does not cover procedures for establishing said postulated relationship.

1.3 This practice serves as an overview of techniques used to verify the applicability of an empirically derived multivariate calibration to the measurement of a sample under test and to verify equivalence between the properties calculated from the empirically derived multivariate calibration and the results of an accepted reference method of measurement to within control limits established for the prespecified statistical confidence level.

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.5 *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.*

<sup>1</sup> This practice is under the jurisdiction of ASTM Committee E13 on Molecular Spectroscopy and Separation Science and is the direct responsibility of Subcommittee E13.11 on Multivariate Analysis.

Current edition approved Dec. 15, 2017. Published February 2018. Originally approved in 2008. Last previous edition approved in 2010 as E2617 – 10. DOI: 10.1520/E2617-17.

## 2. Referenced Documents

2.1 *ASTM Standards:*<sup>2</sup>

- E131 Terminology Relating to Molecular Spectroscopy
- E1655 Practices for Infrared Multivariate Quantitative Analysis
- E1790 Practice for Near Infrared Qualitative Analysis

## 3. Terminology

3.1 For terminology related to molecular spectroscopic methods, refer to Terminology E131. For terminology related to multivariate quantitative modeling refer to Practices E1655. While Practices E1655 is written in the context of multivariate spectroscopic methods, the terminology is also applicable to other multivariate technologies.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *accuracy*—the closeness of agreement between a test result and an accepted reference value.

3.2.2 *bias*—the arithmetic average difference between the reference values and the values produced by the analytical method under test, for a set of samples.

3.2.3 *detection limit*—the lowest level of a property in a sample that can be detected, but not necessarily quantified, by the measurement system.

3.2.4 *estimate*—the constituent concentration, identification, or other property of a sample as determined by the analytical method being validated.

3.2.5 *initial validation*—validation that is performed when an analyzer system is initially installed or after major maintenance.

3.2.6 *Negative Fraction Identified*—the fraction of samples not having a particular characteristic that is identified as not having that characteristic.

<sup>2</sup> 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.

3.2.6.1 *Discussion*—Negative Fraction Identified assumes that the characteristic that the test measures either is or is not present. It is not applicable to tests with multiple possible outcomes.

3.2.7 *ongoing periodic revalidation*—the quality assurance process by which, in the case of quantitative calibrations, the bias and precision or, in the case of qualitative calibrations, the Positive Fraction Identified and Negative Fraction Identified performance determined during initial validation are shown to be sustained.

3.2.8 *Positive Fraction Identified*—the fraction of samples having a particular characteristic that is identified as having that characteristic.

3.2.8.1 *Discussion*—Positive Fraction Identified assumes that the characteristic that the test measures either is or is not present. It is not applicable to tests with multiple possible outcomes.

3.2.9 *precision*—the closeness of agreement between independent test results obtained under stipulated conditions.

3.2.9.1 *Discussion*—Precision may be a measure of either the degree of reproducibility or degree of repeatability of the analytical method under normal operating conditions. In this context, reproducibility refers to the use of the analytical procedure in different laboratories, as in a collaborative study.

3.2.10 *quantification limit*—the lowest level of a sample property which can be determined with acceptable precision and accuracy under the stated experimental conditions.

3.2.11 *range*—the interval between the upper and lower levels of a property (including these levels) that has been demonstrated to be determined with a suitable level of precision and accuracy using the method as specified.

3.2.12 *reference value*—the metric of a property as determined by well-characterized method, the accuracy of which has been stated or defined, that is, another, already-validated method.

3.2.13 *validation*—the statistically quantified judgment that an empirically derived multivariate calibration is applicable to the measurement on which the calibration is to be applied and can perform property estimates with, in the case of quantitative calibrations, acceptable precision, accuracy and bias or, in the case of qualitative calibrations, acceptable Positive Fraction Identified and Negative Fraction Identified, as compared with results from an accepted reference method.

3.2.14 *validation space*—the region(s) of a calibration's multivariate sample space populated by the independent validation samples which are used to validate the calibration.

## 4. Summary of Practice

4.1 Validating an empirically derived multivariate calibration (model) consists of four major procedures: validation at initial development, revalidation at initial deployment or after a revision, ongoing periodic revalidation, and qualification of each measurement before using the calibration to estimate the property(s) of the sample being measured.

## 5. Significance and Use

5.1 This practice outlines a universally applicable procedure to validate the performance of a quantitative or qualitative, empirically derived, multivariate calibration relative to an accepted reference method.

5.2 This practice provides procedures for evaluating the capability of a calibration to provide reliable estimations relative to an accepted reference method.

5.3 This practice provides purchasers of a measurement system that incorporates an empirically derived multivariate calibration with options for specifying validation requirements to ensure that the system is capable of providing estimations with an appropriate degree of agreement with an accepted reference method.

5.4 This practice provides the user of a measurement system that incorporates an empirically derived multivariate calibration with procedures capable of providing information that may be useful for ongoing quality assurance of the performance of the measurement system.

5.5 Validation information obtained in the application of this practice is applicable only to the material type and property range of the materials used to perform the validation and only for the individual measurement system on which the practice is completely applied. It is the user's responsibility to select the property levels and the compositional characteristics of the validation samples such that they are suitable to the application. This practice allows the user to write a comprehensive validation statement for the analyzer system including specific limits for the validated range of application and specific restrictions to the permitted uses of the measurement system. Users are cautioned against extrapolation of validation results beyond the material type(s) and property range(s) used to obtain these results.

5.6 Users are cautioned that a validated empirically derived multivariate calibration is applicable only to samples that fall within the subset population represented in the validation set. The estimation from an empirically derived multivariate calibration can only be validated when the applicability of the calibration is explicitly established for the particular measurement for which the estimation is produced. Applicability cannot be assumed.

## 6. Methods and Considerations

6.1 When validating an empirically derived multivariate calibration, it is the responsibility of the user to describe the measurement system and the required level of agreement between the estimations produced by the calibration and the accepted reference method(s).

6.2 When validating a measurement system incorporating an empirically derived multivariate calibration, it is the responsibility of the user to satisfy the requirements of any applicable tests specific to the measurement system including any Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ) requirements; which may be mandated by competent regulatory authorities, an applicable

Quality Assurance (QA), or Standard Operating Procedure (SOP) or be recommended by the instrument or equipment manufacturer.

### 6.3 Reference Values and Quality Controls for the Accepted Reference Method:

6.3.1 The reference (or true) value which is compared with each respective estimate produced by the empirically derived multivariate calibration is established by applying an accepted reference method, the characteristics of which are known and stated, to the sample from which the measurement system derives the measurement.

6.3.2 To ensure the reliability of the reference values provided by an accepted reference method, appropriate quality controls should be applied to the accepted reference method.

## 7. Procedure

7.1 The objective of the validation procedure is to quantify the performance of an empirically derived multivariate calibration in terms of, in the case of quantitative calibrations, precision, accuracy and bias or, in the case of qualitative calibrations, Positive Fraction Identified and Negative Fraction Identified relative to an accepted reference method for each property of interest. The user must specify, based on the intended use of the calibration, acceptable precision and bias or Positive Fraction Identified and Negative Fraction Identified performance criteria before initiating the validation. These criteria will be dependent on the intended use of the analyzer and may be based, all or in part, on risk based criteria.

7.1.1 The acceptable performance criteria specified by the user may be constant over the entire range of sample variability. Alternatively, different acceptable performance criteria may be specified by the user for different sub-ranges of the full sample variability.

7.2 Validation of calibration is accomplished by using the calibration to estimate the property(s) of a set of validation samples and statistically comparing the estimates for these samples to known reference values. Validation requires thorough testing of the model with a sufficient number of representative validation samples to ensure that it performs adequately over the entire range of possible sample variability.

### 7.3 Initial Validation Sample Set:

7.3.1 For the initial validation of a multivariate model, an ideal validation sample set will:

7.3.1.1 Contain samples that provide sufficient examples of all combinations of variation in the sample properties which are expected to be present in the samples which are to be analyzed using the calibration;

7.3.1.2 Contain samples for which the ranges of variation in the sample properties is comparable to the ranges of variation expected for samples that are to be analyzed using the model;

7.3.1.3 Contain samples for which the respective variations of the sample properties are uniformly and mutually independently distributed over their full respective ranges or, when applicable, subranges of variation; and

7.3.1.4 Contain a sufficient number of samples to statistically test the relationships between the measured variables and the properties that are modeled by the calibration.

7.3.2 For simple systems, sufficient validation samples can generally be obtained to meet the criteria in 7.3.1.1 – 7.3.1.4. For complex mixtures, obtaining an ideal validation set may be difficult if not impossible. In such cases, it may be necessary to validate discrete subranges of the calibration incrementally, over time as samples become available.

7.3.3 The number of samples needed to validate a calibration depends on the complexity of the calibration, the ranges of property variation over which the calibration is to be applied, and the degree of confidence required. It is important to validate a calibration with as many samples as possible to maximize the likelihood of challenging the calibration with rarely occurring, but potentially troublesome samples. The number and range of validation samples should be sufficient to validate the calibration to the statistical degree of confidence required for the application. In all cases, a minimum of 20 validation samples is recommended. In addition, the validation samples should:

7.3.3.1 Multivariately span the ranges of sample property values over which the calibration will be used; that is, the span and the standard deviation of the ranges of sample property values for the validation samples should be at least 100 % of the spans of the sample property values over which the calibration will be used, and the sample property values for the validation samples should be distributed as uniformly as possible throughout their respective ranges, and the variations of the sample property values among the samples should be as mutually independent as possible; and

7.3.3.2 Span the ranges of the independent variables over which the calibration will be used; that is, if the range of an independent variable is expected to vary from  $a$  to  $b$ , and the standard deviation of the independent variable is  $c$ , then the variations of that independent variable in the set of validation samples should cover at least 100 % of the range from  $a$  to  $b$ , and should be distributed as uniformly as possible across the range such that the standard deviation in that independent variable estimated for the validation samples will be at least 95 % of  $c$ .

(1) When validating a calibration for which detection limit or quantification limit is an important consideration, the user should include a number of validation samples whose property(s) are close to the detection or quantification limit(s) sufficient to validate the respective limit(s) to the statistical degree of confidence required for the application.

7.4 For quantitative calibrations, the validation error for each property in each sample is given by the Standard Error of Validation (SEV) and bias for that property.

7.4.1 The validation bias,  $ev$ , is a measure of the average difference between the estimates made based on the empirical model and the results obtained on the same validation samples using the reference method.

7.4.1.1 If there are single reference values and estimates for each validation sample, the validation bias is calculated as:

$$\bar{e}_v = \frac{\sum_{i=1}^v (\hat{v}_i - v_i)}{v} \quad (1)$$