



Standard Practice for Evaluating Equivalence of Two Testing Processes¹

This standard is issued under the fixed designation E2935; 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 provides statistical methodology for conducting equivalence studies on numerical data from two sources of test results to determine if their true means, variances, or other parameters differ by no more than predetermined limits.

1.2 Applications include (1) equivalence studies for bias against an accepted reference value, (2) determining means equivalence of two test methods, test apparatus, instruments, reagent sources, or operators within a laboratory or equivalence of two laboratories in a method transfer, and (3) determining non-inferiority of a modified test procedure versus a current test procedure with respect to a performance characteristic.

1.3 The guidance in this standard applies to experiments conducted either on a single material at a given level of the test result or on multiple materials covering a selected range of test results.

1.4 Guidance is given for determining the amount of data required for an equivalence study. The control of risks associated with the equivalence decision is discussed.

1.5 The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.

1.6 *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.*

1.7 *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.*

¹ This test method is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.20 on Test Method Evaluation and Quality Control.

Current edition approved June 1, 2021. Published June 2021. Originally approved in 2013. Last previous edition approved in 2020 as E2935 – 20^ε. DOI: 10.1520/E2935-21.

2. Referenced Documents

2.1 ASTM Standards:²

E122 Practice for Calculating Sample Size to Estimate, With Specified Precision, the Average for a Characteristic of a Lot or Process

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

E456 Terminology Relating to Quality and Statistics

E2282 Guide for Defining the Test Result of a Test Method

E2586 Practice for Calculating and Using Basic Statistics

E3080 Practice for Regression Analysis with a Single Predictor Variable

2.2 USP Standard:³

USP <1223> Validation of Alternative Microbiological Methods

3. Terminology

3.1 *Definitions*—See Terminology E456 for a more extensive listing of statistical terms.

3.1.1 *accepted reference value, n*—a value that serves as an agreed-upon reference for comparison, and which is derived as: (1) a theoretical or established value, based on scientific principles, (2) an assigned or certified value, based on experimental work of some national or international organization, or (3) a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group.

E177

3.1.2 *bias, n*—the difference between the expectation of the test results and an accepted reference value.

E177

3.1.3 *confidence interval, n*—an interval estimate [L, U] with the statistics L and U as limits for the parameter θ and with confidence level $1 - \alpha$, where $\Pr(L \leq \theta \leq U) \geq 1 - \alpha$.

E2586

3.1.3.1 *Discussion*—The confidence level, $1 - \alpha$, reflects the proportion of cases that the confidence interval [L, U] would contain or cover the true parameter value in a series of repeated random samples under identical conditions. Once L and U are

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.

³ Available from U.S. Pharmacopeial Convention (USP), 12601 Twinbrook Pkwy., Rockville, MD 20852-1790, <http://www.usp.org>.

given values, the resulting confidence interval either does or does not contain it. In this sense “confidence” applies not to the particular interval but only to the long run proportion of cases when repeating the procedure many times.

3.1.4 *confidence level*, n —the value, $1 - \alpha$, of the probability associated with a confidence interval, often expressed as a percentage. **E2586**

3.1.4.1 *Discussion*— α is generally a small number. Confidence level is often 95 % or 99 %.

3.1.5 *confidence limit*, n —each of the limits, L and U, of a confidence interval, or the limit of a one-sided confidence interval. **E2586**

3.1.6 *degrees of freedom*, n —the number of independent data points minus the number of parameters that have to be estimated before calculating the variance. **E2586**

3.1.7 *equivalence*, n —condition that two population parameters differ by no more than predetermined limits.

3.1.8 *intermediate precision conditions*, n —conditions under which test results are obtained with the same test method using test units or test specimens taken at random from a single quantity of material that is as nearly homogeneous as possible, and with changing conditions such as operator, measuring equipment, location within the laboratory, and time. **E177**

3.1.9 *mean*, n —of a population, μ , average or expected value of a characteristic in a population; of a sample, $\bar{X}$ sum of the observed values in the sample divided by the sample size. **E2586**

3.1.10 *percentile*, n —quantile of a sample or a population, for which the fraction less than or equal to the value is expressed as a percentage. **E2586**

3.1.11 *population*, n —the totality of items or units of material under consideration. **E2586**

3.1.12 *population parameter*, n —summary measure of the values of some characteristic of a population. **E2586**

3.1.13 *precision*, n —the closeness of agreement between independent test results obtained under stipulated conditions. **E177**

3.1.14 *quantile*, n —value such that a fraction f of the sample or population is less than or equal to that value. **E2586**

3.1.15 *repeatability*, n —precision of test results from tests conducted within the shortest practical time period on identical material by the same *test method* in a single laboratory with all known sources of variability conditions controlled at the same levels (see **repeatability conditions**). **E177**

3.1.16 *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. **E177**

3.1.17 *repeatability standard deviation* (s_r), n —the standard deviation of test results obtained under repeatability conditions. **E177**

3.1.18 *sample*, n —a group of observations or test results, taken from a larger collection of observations or test results,

which serves to provide information that may be used as a basis for making a decision concerning the larger collection. **E2586**

3.1.19 *sample size*, n , n —number of observed values in the sample. **E2586**

3.1.20 *sample statistic*, n —summary measure of the observed values of a sample. **E2586**

3.1.21 *standard deviation—of a population*, σ , the square root of the average or expected value of the squared deviation of a variable from its mean; *of a sample*, s , the square root of the sum of the squared deviations of the observed values in the sample from their mean divided by the sample size minus 1. **E2586**

3.1.22 *test result*, n —the value of a characteristic obtained by carrying out a specified test method. **E2282**

3.1.23 *test unit*, n —the total quantity of material (containing one or more test specimens) needed to obtain a test result as specified in the test method. See test result. **E2282**

3.1.24 *variance*, σ^2 , s^2 , n —square of the standard deviation of the population or sample. **E2586**

3.2 Definitions of Terms Specific to This Standard:

3.2.1 *bias equivalence*, n —equivalence of a population mean with an accepted reference value.

3.2.2 *equivalence limit*, E , n —in *equivalence testing*, a limit on the difference between two population parameters.

3.2.2.1 *Discussion*—In certain applications, this may be termed *practical limit* or *practical difference*.

3.2.3 *equivalence test*, n —a statistical test conducted within predetermined risks to confirm equivalence of two population parameters.

3.2.4 *means equivalence*, n —equivalence of two population means.

3.2.5 *non-inferiority*, n —condition that the difference in means or variances of test results between a modified testing process and a current testing process with respect to a performance characteristic is no greater than a predetermined limit in the direction of inferiority of the modified process to the current process.

3.2.5.1 *Discussion*—Other terms used for *non-inferior* are “equivalent or better” or “at least equivalent as.”

3.2.6 *paired samples design*, n —in *means equivalence testing*, single samples are taken from the two populations at a number of sampling points.

3.2.7 *power*, n —in *equivalence testing*, the probability of accepting equivalence, given the true difference between two population means.

3.2.7.1 *Discussion*—In the case of testing for bias equivalence the power is the probability of accepting equivalence, given the true difference between a population mean and an accepted reference value.

3.2.8 *range equivalence*, n —equivalence of two population means over a range of test result values.

3.2.9 *slope equivalence*, n —equivalence of the slope of a linear statistical relationship with the value one (1).