



Standard Practice for Statistical Assessment and Improvement of Expected Agreement Between Two Test Methods that Purport to Measure the Same Property of a Material¹

This standard is issued under the fixed designation D6708; 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 statistical methodology for assessing the expected agreement between two different standard test methods that purport to measure the same property of a material, and for the purpose of deciding if a simple linear bias correction can further improve the expected agreement. It is intended for use with results obtained from interlaboratory studies meeting the requirement of Practice D6300 or equivalent (for example, ISO 4259). The interlaboratory studies shall be conducted on at least ten materials in common that among them span the intersecting scopes of the test methods, and results shall be obtained from at least six laboratories using each method. Requirements in this practice shall be met in order for the assessment to be considered suitable for publication in either method, if such publication includes claim to have been carried out in compliance with this practice. Any such publication shall include mandatory information regarding certain details of the assessment outcome as specified in the Report section of this practice.

1.2 The statistical methodology is based on the premise that a bias correction will not be needed. In the absence of strong statistical evidence that a bias correction would result in better agreement between the two methods, a bias correction is not made. If a bias correction is required, then the *parsimony principle* is followed whereby a simple correction is to be favored over a more complex one.

NOTE 1—Failure to adhere to the parsimony principle generally results in models that are over-fitted and do not perform well in practice.

1.3 The bias corrections of this practice are limited to a constant correction, proportional correction, or a linear (proportional + constant) correction.

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 May 2021. Originally approved in 2001. Last previous edition approved in 2019 as D6708 – 19a^{ε1}. DOI: 10.1520/D6708-21.

1.4 The bias-correction methods of this practice are method symmetric, in the sense that equivalent corrections are obtained regardless of which method is bias-corrected to match the other.

1.5 A methodology is presented for establishing the numerical limit (designated by this practice as the *between methods reproducibility*) that would be exceeded about 5 % of the time (one case in 20 in the long run) for the difference between two results where each result is obtained by a different operator using different apparatus and each applying one of the two methods *X* and *Y* on identical material, where one of the methods has been appropriately bias-corrected in accordance with this practice, in the normal and correct operation of both test methods.

NOTE 2—In earlier versions of this standard practice, the term “cross-method reproducibility” was used in place of the term “between methods reproducibility.” The change was made because the “between methods reproducibility” term is more intuitive and less confusing. It is important to note that these two terms are synonymous and interchangeable with one another, especially in cases where the “cross-method reproducibility” term was subsequently referenced by name in methods where a D6708 assessment was performed, before the change in terminology in this standard practice was adopted.

NOTE 3—Users are cautioned against applying the between methods reproducibility as calculated from this practice to materials that are significantly different in composition from those actually studied, as the ability of this practice to detect and address sample-specific biases (see 6.7) is dependent on the materials selected for the interlaboratory study. When sample-specific biases are present, the types and ranges of samples may need to be expanded significantly from the minimum of ten as specified in this practice in order to obtain a more comprehensive and reliable between methods reproducibility that adequately cover the range of sample-specific biases for different types of materials.

1.6 This practice is intended for test methods which measure quantitative (numerical) properties of petroleum or petroleum products.

1.7 The statistical calculations of this practice are also applicable for assessing the expected agreement between two different test methods that purport to measure the same property of a material using results that are not as described in 1.1, provided the results and associated statistics from each test method are obtained from a specifically designed multi-lab study or from a proficiency testing program (e.g.: ILCP) where

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