Validation and Verification of Analytical Testing Methods Used for Tobacco Products

---

Guidance for Industry

This draft guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft guidance should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to http://www.regulations.gov. Submit written comments to the Dockets Management Staff, Food and Drug Administration, 5630 Fishers Lane, Room 1061, Rockville, MD, 20852. All comments should be identified with the docket number FDA-2021-D-0756.

For questions regarding this draft guidance, contact the Center for Tobacco Products at (Tel) 1-877-CTP-1373 (1-877-287-1373) Monday-Friday, 9 a.m. – 4 p.m. EDT.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Tobacco Products

December 2021

Table of Contents

Validation and Verification of Analytical Testing Methods Used for Tobacco Products

Guidance for Industry[1]

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. INTRODUCTION

FDA is issuing this guidance to provide information and recommendations on how tobacco product manufacturers can produce validation and verification data for the analytical procedures and methods used to support regulatory submissions for finished tobacco products, such as substantial equivalence applications (SE), premarket tobacco product applications (PMTA), and modified risk tobacco product applications (MRTPA). The recommendations address analytical testing of tobacco product constituents, ingredients, and additives, as well as stability testing of finished tobacco products. The principles in this guidance may also be useful for finished tobacco product testing and reporting of harmful and potentially harmful constituents (HPHCs) in tobacco products and tobacco smoke. Applications often have data and information to support analytical methods used for testing, and this guidance is intended to help tobacco product manufacturers and analytical laboratories assemble and present scientifically valid data and information to support the analytical methodologies used for regulatory submissions of finished tobacco products. The recommendations in this guidance will help FDA better understand and assess the reliability of the data presented in regulatory submissions. The use of the words *acceptable* and *acceptability* in this guidance refer to the relative reliability of test results; they do not mean that CTP will or will not accept the study or regulatory submission.

The contents of this document do not have the force and effect of law and are not meant to bind the public in any way, unless specifically incorporated into a contract. This document is intended only to provide clarity to the public regarding existing requirements under the law. FDA guidance documents, including this guidance, should be viewed only as recommendations, unless specific

[1] This guidance was prepared by the Office of Science, Office of Enforcement, and Office of Regulations in the Center for Tobacco Products at FDA.

regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. BACKGROUND

The Family Smoking Prevention and Tobacco Control Act (Pub. L. 111–31) (Tobacco Control Act), enacted on June 22, 2009, amended the Federal Food, Drug, and Cosmetic Act (FD&C Act) and provided FDA with the authority to regulate the manufacture, marketing, and distribution of cigarettes, cigarette tobacco, roll-your-own (RYO) tobacco, smokeless tobacco products to protect the public health and to reduce tobacco use by minors. The Tobacco Control Act additionally gave FDA authority to issue regulations deeming other tobacco products to be subject to FDA's authority. In May of 2016, FDA issued a final rule deeming all products that meet the statutory definition of "tobacco product" in the FD&C Act, except accessories of such tobacco products, to be subject to Chapter IX of the FD&C Act.

The FD&C Act requires, among other things, premarket review for new tobacco products and modified risk tobacco products, [*see* Sections 910 and 911 of the FD&C Act], and also reporting of harmful and potentially harmful constituents under section 904. Information about constituents, for example, might be required by law or otherwise support the findings for premarket authorization. Regulatory submissions often contain data from analytical testing, such as data about ingredients, constituents, and additives. In standard practice, analytical testing is done through the use of validated analytical methods. In these cases, the applicant will want to use analytical methods that are sufficiently precise, accurate, selective, sensitive and have suitable linearity and range. Validation involves documenting, through the use of specific laboratory investigations, that the performance characteristics of the method, including the analytical procedure and sampling procedure, are suitable and reliable for the intended analytical applications, in terms of precision, accuracy, selectivity, and sensitivity.

III. DEFINITIONS

For purposes of this guidance only, the following definitions apply:

Accuracy of an analytical method describes the closeness of mean test results obtained by the analytical method to the true value (concentration or mass) of the analyte. The accuracy of the method approximates the determinate error in a measurement.

Analytical procedure is a detailed step-by-step instruction that describes sample preparation from stock solution or homogeneous sample through to the instrument measurement of the samples and standards.

Analytical sampling procedure (Sampling procedure) describes a detailed step-by-step sampling of the tobacco product for the analyte(s) of interest. This procedure describes all of the steps necessary to prepare a stock solution or homogeneous sample suitable for the intended analytical

procedure. [Note: this is different from the *Sampling protocol* which describes the steps taken to obtain a representative sample of the tobacco product under test. The sampling protocol is not further described in this document.]

Analytical test method consists of the analytical sampling procedure and analytical measurement procedure employed by the applicant.

Homogenous sample is a portion of the material under test that is internally consistent and representative of the bulk of the material.

Intermediate precision is a measurement of the degree of variability in a given measurement or set of measurements within a laboratory which may be typical of expected changes in the laboratory. Changes that are expected to occur include temporal changes (e.g., across different days), collection of data by different individuals, and collection using different instruments and sampling setups.

Limit of detection is the lowest analyte concentration that can be reliably differentiated from the background noise.

Limit of quantification is the lowest analyte concentration that can be quantified by the analytical test method with acceptable precision and accuracy.

Linearity is the ability of a method, within a certain range, to provide an instrumental response or test results proportional to the quantity of the analyte in the test sample.

Performance acceptance criteria (Performance criteria) are the pre-determined requirements, defined by the analyst, that describe acceptable values for the validation or verification parameters of the method or procedure subject to examination.

Precision of an analytical method describes the closeness of individual measures of an analyte when a procedure is applied repeatedly to multiple aliquots of a single homogeneous solution of an analyte. Precision approximates the indeterminate error in a measurement, in that a greater precision of a measurement denotes a lower indeterminate error of that measurement. Precision is a combination of repeatability, intermediate precision, reproducibility, and robustness. [Note: reproducibility is not a subject of this guidance because it is based on interlaboratory variability]

Range is the interval of analyte concentrations over which the method provides suitable accuracy and precision.

Repeatability is precision obtained under observation conditions where independent test results are obtained with the same method on identical test items in the same test facility by the same operator using the same equipment within short intervals of time.

Resolution is the chromatographic separation of two components in a mixture, calculated by:

$$R_s = 1.18 \times (t_{R2} - t_{R1})/(W_{1,h/2} + W_{2,h/2}),$$

where t_{R2} and t_{R1} are the retention times of the two components; and $W_{1,h/2}$ and $W_{2,h/2}$ are the corresponding widths at the bases of the peaks obtained at half the height.

Robustness is the measure of an analytical test method's capacity to remain unaffected by small, but deliberate, variations in method parameters and provides an indication of the reliability of the method during normal usage.

Sample matrix (Matrix) is everything that is part of the sample except for the analyte(s) of interest.

Selectivity is the ability of an analytical method to differentiate and quantify the analyte of interest in the presence of other matrix components present in the sample. Selectivity should be established at the limit of quantification.

Sensitivity is the change in response of the test method with regard to a change in concentration of the analyte.

Stock solution is a concentrated solution that may be diluted to lower concentrations for use in testing.

Tobacco product reference standard is a tobacco product with known characteristics used for quality control and quality assurance during preparation and analysis of test material.

Total Error Probability is a statistical approach of combining the accuracy error and precision error and is considered in terms of a probability function. This probability function demonstrates that, given these errors, the likelihood that the true value of the analyte can be determined is within a user-defined confidence interval.

Validation means a process of demonstrating or confirming that the analytical test method is suitable and reliable for its intended purpose. The suitability and reliability of a method is determined by comparing the results of measurements of the validation parameters to the performance criteria of the intended purpose. Validation of an analytical method applies to a specific laboratory, for a specific tobacco product formulation, and equipment performing the analytical test method for an intended use over a reasonable period of time. A validated method can be extended to other tobacco product formulations, different laboratories, and across minor changes in equipment (similar to acceptable adjustments described in *USP-NF Chromatography <621>*) through a verification study.

Validation parameters include accuracy (recovery), precision (repeatability, intermediate precision, and robustness), linearity/range, selectivity, and sensitivity (limit of quantification and limit of detection).

Verification means the demonstration of a laboratory's ability to successfully meet performance criteria established for an analytical test method previously validated in the laboratory performing the verification. Verification is typically recommended following a change to one of the procedures in a method or a change to the tobacco product being tested. The extent of verification studies needed varies depending upon the extent of the change(s) to the method, and may include accuracy and precision, selectivity, limit of quantification, or, as appropriate, other critical validation parameters.

IV. TOBACCO PRODUCT ANALYTICAL TEST METHODS

A. General Recommendations

FDA recommends tobacco product manufacturers consider using testing facilities that are accredited and compliant with the standards listed in this guidance when conducting validation and verification of analytical test methods used to support a reporting submission, record, or document. Use of these facilities provides greater confidence in the consistency of the results that may be expected from that facility. It is recommended that records of accreditation and compliance be maintained in case they are requested by FDA[2]. These records are useful for checking the veracity of data and may be needed for some FDA submissions (e.g., premarket applications).

At the most basic level, validation of a procedure is a systematic evaluation of the errors inherent in the measurement of a sample. A procedure may be determined to be suitable for its intended purpose when the calculated total error falls below the acceptable error of the measurement. A procedure may be determined to be reliable though the evaluation of the total error relative to acceptable error over a statistically valid period of time. The validation procedure described in this guidance provides a means to demonstrate the suitability and reliability of an analytical test method.

Validation of an analytical test method includes two separate major components, the analytical procedure validation and the sampling procedure validation. The two major components may be modular in nature. The analytical procedure may be used for a number of analytes and may be applied to different tobacco product matrices. Likewise, the sampling procedure validation may also be applied to different analytes and different tobacco product categories and sub-categories.

The analytical procedure validation is intended to define the performance limitations of the analytical procedure. The performance limitations are defined by the performance criteria of the procedure, which are defined prior to the beginning of method development and the subsequent validation. The performance criteria typically include information such as the target analyte, an approximation of the range of concentrations of the analyte in the sample, the intended purpose of the procedure (e.g., qualitative, quantitative, major component, minor component, etc.), and the number of samples to be analyzed. Performance criteria are typically described through validation parameters with associated acceptance criteria. The validation parameters that are

[2] FD&C Act sections 904, 909, 910, 911, 915.

recommended for a validation study of a quantitative analytical procedure include: accuracy, precision, selectivity, sensitivity, linearity, and range. These parameters help ensure that the procedures used to report constituents, ingredients, additives, and stability testing are suitable, accurate, precise enough to produce reliable results over time, and that they are sensitive and selective enough to ensure that the reported values are representative of the actual product under test. Results obtained from a validated procedure demonstrate that the characteristics reported are representative of the actual product under test, and the calculation of exposure and risks associated with those characteristics can be appropriately considered. An example of performance criteria for an analytical method might be as described by Table 1.

Table 1: Hypothetical Example Performance Criteria for Nicotine Assay of Tobacco Filler[3]

Validation Parameter(s)	Measurement Approach	Acceptance Criteria
Accuracy and Repeatability	Lower Limit: 85% Upper Limit: 115%	Total Error Probability[4] Not Less Than 95.0%
Accuracy and Intermediate Precision	Lower Limit: 85% Upper Limit: 115%	Total Error Probability Not Less Than 95.0%
Validation Parameter(s)	Measurement Approach	Acceptance Criteria
Robustness	Lower Limit: 85% Upper Limit: 115%	Total Error Probability Not Less Than 95.0%
Selectivity	Resolution	Not Less Than 1.5
Limit of Quantification	Serial Dilution (µg/g)	Not More Than 0.1
Limit of Detection	S/N approximation (µg/g)	Not More Than 0.05
Linearity	Total Error Probability at Each Point (5 concentrations, 7 replicates at each concentration) Lower Limit: 85% Upper Limit: 115%	Total Error Probability Not Less than 95%

Measurements for each analyte in the matrix should be independently validated. Analytical procedure validations are conducted using standard materials in compatible solutions (see Section V). Note that the analytical procedure encompasses all of the sample manipulation steps (typically only dilution) beginning from the stock sample solution.

The sampling procedure validation is intended to define the performance limitations of the sample collection and manipulation necessary to obtain a stock sample solution. Performance criteria of the sampling procedure are also described through validation parameters with associated acceptance criteria. The validation parameters that are recommended for a validation

[3] Note: this table uses acceptance criteria based on the approaches described more fully in the remainder of this guidance.

[4] See section IV(A)(1)(a).

study of a quantitative sampling procedure include: accuracy, precision, sensitivity, linearity, and range.

Throughout this guidance, a recommended minimum number of replicates for a given validation parameter is provided. These recommendations are generally recognized within the scientific community[5] as the number of replicates to substantiate the results of the type of testing conducted but more replicates will generally provide better estimations of the true value being measured. The reliability of the results of the validation are directly affected by the number of replicates collected. The number of replicates necessary to obtain an acceptable probability for validation purposes does not affect the number of replicates that should be used during the routine usage of the method or procedure. However, in cases where the material or the method is more variable than the acceptance ranges, increasing the number of replicates is a general practice and is recommended. Given the wide array of methods and procedure that can be validated, it is important to tailor the number of replicates used for validation purposes to the method or procedure being validated. Some method validations may be performed with fewer replicates than what is recommended in this guidance, while others may need a larger number of replicates to meet the performance criteria established for the method or procedure being validated.

1. Accuracy and Precision

The accuracy of an analytical method describes the closeness of mean test results obtained by the method to the true value (concentration or mass) of the analyte. The accuracy approximates the determinate error in a measurement. Accuracy is determined by replicate analysis of samples containing known amounts of the analyte. Measuring accuracy generally means that several tests should be done at any given concentration of a known amount of analyte. Generally, seven or more determinations per concentration can provide sufficient data to evaluate accuracy.
In general, the more replicates used the better the statistical analysis and, therefore, the better the subsequent determination of variability of the assay to determine a specific amount of an analyte. More replicates provides more data and the greater amount of data allows more accuracy and certainty in the conclusion.

For accuracy, as opposed to precision, more replicates are needed over a range of concentrations of analyte because of the nature of the error being determined. Accuracy measures errors that are always present in every measurement. Every measurement also contains random error that changes with each measurement. Random errors can be either positive or negative, and therefore a larger collection of values tends to cause these errors to cancel themselves out of the calculation, allowing the underlying accuracy to be determined. Therefore, more replicates

[5] *See generally:* DeBievre, P. and H. Gunzler (2005). Validation in Chemical Measurement, Springer Science & Business Media (https://doi.org/10.1007/b138530) and;
Mark T. Stauffer (2018). Calibration and Validation of Analytical Methods - A Sampling of Current Approaches, IntechOpen Publishing (https://doi.org/10.5772/intechopen.69918) and;
Marc Andre Goulet and Denis Cousineau (2019). The Power of Replicated Measures to Increase Statistical Power, Advances in Methods and Practices in Psychological Science (https://doi.org/10.1177/2515245919849434).

should generally be used to calculate accuracy than what is used to calculate precision, or linearity.

The precision of an analytical method describes the closeness of individual measures of an analyte when the procedure is applied repeatedly to multiple aliquots of a single homogeneous solution of an analyte. Precision approximates the indeterminate error in a measurement. The precision value is calculated as the coefficient of variation (CV) of the measurement and is approximately the same as the expected analyte concentration in the intended sample (referred to as "target concentration" hereafter). The CV is often referred to as the percent relative standard deviation (%RSD) and the terms are used interchangeably in most laboratories. Precision can be further subdivided into repeatability (ISO-r), reproducibility (ISO-R), and intermediate precision, each of which is evaluated differently depending upon whether the analytical procedure, sampling procedure, or analytical method is being examined. These concepts are discussed in greater detail in sections V, VI, and VII, respectively. Generally, five determinations or more per concentration can provide sufficient data to evaluate precision. In the case of a method with high variability, additional replicates are beneficial and are recommended.

Because every measurement is a summation of the true value, the determinate error, and the indeterminate error, knowledge of all errors is important for an approximation of the true value. Thus, a special effort should be made to isolate and quantify the relative amounts of the two error types from the true value. This is generally accomplished by measuring replicate solutions of a reference standard material at a known concentration. Depending upon which portion of the validation is being considered (e.g., accuracy, precision), the standard solutions are treated differently, and the number of replicate solutions will change. This is discussed further in sections V, VI, and VII.

Consider an example in which an analyst made two sets of replicate solutions using reference standard materials. One solution set with a target concentration of 1.0 µg/mL and the other set with a target concentration of 10 µg/mL. The analyst then made the measurements shown in Table 2.

Table 2: Hypothetical Example test results from a method with different concentrations of nicotine reference standard[6]

Replicate	Calculated concentration (µg/mL) (Target: 1 µg/mL)	Calculated concentration (µg/mL) (Target: 10 µg/mL)
1	1.1	10.5
2	0.8	10.5
3	1.3	10.4
4	1.0	10.2
5	0.7	10.6
6	0.7	10.1
7	1.0	10.0
Mean	0.94	10.32

[6] Note that these data are for example purposes only and do not represent any particular tobacco product.

% Recovery	94%	103.2%
St. Dev.	0.22	0.22

Based on these measurements, the accuracy of the 1 µg/mL set would be calculated as a % recovery of 94% and precision would be calculated as 22.2% RSD. The 10 µg/mL set would have a % recovery of 103.2% and a precision of 2.2% RSD. Because it is unclear if either or both of these data sets would have an acceptable total error of measurement, the total error of the measurement should be calculated using the formula below.

a. Total Error of a Measurement

The accuracy and precision values may be considered individually or in a total error context. Typically, analytical procedures represent a compromise and balance among sensitivity, accuracy, and precision. Therefore, setting acceptance criteria based upon consideration of precision and accuracy in isolation may result in a validation expectation that cannot be met. Instead, a total error evaluation provides a better means to demonstrate the suitability of procedures and methods for their intended purposes. A total error calculation is not a simple summation of the precision and accuracy values. Instead, total error is considered in terms of a probability function. This probability function is the likelihood that the true value of an analyte at the target concentration will pass the acceptance criteria of the analyte at a user defined confidence interval.[7] FDA recommends the use of a 95% confidence interval as this is the scientifically valid standard. This probability is calculated as:

Total Error Probability = $1- t(q_U,df) - t(q_L,df)$,
where
$t(q,df)$ = cumulative Student's t distribution with degrees of freedom (df) evaluated at quantile, q
q_U *= (Upper-Mean) / (RSD*√(1 + 1/n))*
q_L *= (Lower-Mean)) / (RSD*√(1 + 1/n))*
$df = n - 1$
n = number of replicate Standard solutions prepared
Upper = upper limit of the generally accepted range of the analyte (e.g., 102.0%)
Lower = lower limit of the generally accepted range of the analyte (e.g., 98.0%)
*Mean = % recovery = average measured value/known value*100%*
*RSD = Standard deviation of a measurement/value of the measurement*100%*

When calculated in this manner, a total error probability of 0.95 or greater indicates the precision and accuracy are suitable for their intended purposes.[8] The selection of suitable upper and lower acceptance criteria of the measurement are critical decisions and are discussed more fully in sections V, VI, and VII.

[7] Ermer, J. and P. W. Nethercote (2014). Method Validation in Pharmaceutical Analysis. Singapore, Wiley-VCH.
[8] One means to obtain this calculation might be to use the following code in Microsoft™ Excel: =1-TDIST(-((Upper-Mean)/(SQRT(1+1/number of reps)*RSD))-TDIST(-((Lower-Mean)/(SQRT(1+1/number of reps)*RSD))

As an example of the application of total error probability, consider the application of the performance criteria included in Table 1 to the data collected in Table 2, assuming that this data is a repeatability experiment. Table 1 states that the upper and lower values are 115 and 85, respectively, with an acceptance criteria of a total error probability not less than 95%. The mean value of the first data set (0.94 µg/mL) is calculated as a percentage of known value, which is 94%, the RSD is calculated as 22.2%, and the number of replicates is seven. The resulting total probability error would be 37%, which does not meet the stated acceptance criteria. Therefore, this method would be considered unsuitable for its intended application (the evaluation of lower concentrations of nicotine). However, the data from the second set would result in a mean value of 103.2% and the RSD is 2.2% RSD, which would yield a total error probability of 99%. This meets the performance criteria and therefore demonstrates that the method may be suitable (pending the evaluation of the other performance criteria) for the evaluation of higher concentrations of nicotine.

2. Selectivity

Selectivity is the ability of an analytical method to differentiate and quantify the analyte in the presence of other components in a sample. Selectivity is evaluated depending upon the type of procedure being validated. For chromatographic procedures, which make up the majority of procedures used in tobacco chemistry, the evaluation of the resolution is made between the analyte and interferences, including background disturbances, enantiomers, similar chemical species, and unidentified co-eluting species. If the method is intended to quantify more than one analyte at a time in the sample, each analyte should be tested to ensure that there is no interference caused by the analytes and other chromatographic features. A resolution of greater than 1.5 is generally considered to be acceptable by chromatographers. However, the lack of chromatographic resolution does not necessarily indicate that a method fails validation, especially if the analytical procedure uses a sufficiently specific detector (like a mass spectrometer, multichannel UV detector, serial detection techniques (e.g., LC-UV-MS)). When chromatographic resolution cannot be or is not achieved, FDA recommends other approaches be considered. One possible alternative approach is the evaluation of test mixtures of important constituents with known concentrations for bias caused by the interference. The additional bias found in these experiments is added to the accuracy value and the total error probability of the method is calculated.

3. Linearity and Range

The linearity and range of a method or procedure in the validation experiment is defined as the range of concentrations for which an acceptable level of precision and accuracy can be achieved. The acceptability of the measurements at each concentration are evaluated using the total error probability described above. The acceptance criteria are specific to the procedure and method being validated and are discussed further in section VII.

A calibration (standard) curve is the relationship between instrument response and known concentrations of the analyte. This type of curve is typically plotted during the method development phase and is then regularly plotted as a component of system suitability evaluations

on an on-going basis. This helps to assure that the test method is functioning as validated and reduces loss of samples or time because of problems found after testing is completed. FDA recommends calibration curves be generated for each analyte in the sample and include at least five concentrations spread across the intended calibration range of the method, with three concentrations in the linear portion of the curve and two approaching the asymptotic portions. During method development, it is further recommended that a sufficient number of samples at each concentration be used to adequately define the relationship between concentration and response. However, calibration curves typically employed for routine use are less rigorous and lack the accepted international guidelines for the scientific stringency necessary for the evaluation of linearity and range components of the validation activity.[9]

As part of validation, you should use the number of replicates that would support the accuracy of the procedure across the range of analyte concentrations. Within the scientific community, it is generally accepted that no fewer than five replicates of a defined concentration of analyte solution at no fewer than five concentrations are to be used. FDA recommends this standard approach. The concentrations selected should span the entire working range of the procedure or method. The range of concentrations should include the limit of quantification, the target concentration level, the highest concentration to which the method/procedure will likely be used, and at least two additional concentrations. The minimum of 25 measured values (5 replicates of 5 concentrations) would be used to calculate a best fit line and the R^2, slope, and intercept should be reported.

a. Limit of Quantification (LOQ)

The limit of quantification is the lowest concentration for which replicate solutions can be measured with acceptable precision and accuracy. The acceptability of the concentrations is evaluated using the total error probability described above. The acceptance criteria are specific to the procedure and method being validated and are discussed further in section VII. The LOQ approximation using a signal to noise ratio may be used as a means to approximate the starting point for LOQ determination but may be inaccurate and lead to the use of procedures that are not suitable for their intended purpose.

b. Limit of Detection (LOD)

The limit of detection is the lowest concentration at which an analyte can be reliably differentiated from the background noise. The traditional approach is approximating this using three times the signal to noise ratio. The LOD can also be empirically determined through serial dilutions of a standard or spiked analyte solution until a predetermined total error probability is found, but the outcomes of this approach are not generally better than the traditional signal to noise approach.

[9] For additional information, see Linearity Of Calibration Curves For Analytical Methods: A Review Of Criteria For Assessment Of Method Reliability by Seyed Mojtaba Moosavi and Sussan Ghassabian, *available at* https://www.intechopen.com/books/calibration-and-validation-of-analytical-methods-a-sampling-of-current-approaches/linearity-of-calibration-curves-for-analytical-methods-a-review-of-criteria-for-assessment-of-method

B. Analytical test method development

Analytical test method development should (1) satisfy the objectives of the analytical test method and (2) meet method performance criteria. All analytical test methods are developed to measure the specified analytes over a specific concentration range. The method development and establishment phase generally define the limits of the linear dynamic range (linearity and sensitivity), selectivity, robustness of the method, and the solution stability and storage conditions. It is important to note that method development and method validation are different activities. The method development may provide an insight into the boundary conditions of a procedure; however, only the evaluation as part of a validation will demonstrate the suitability of the procedure for the intended purpose. Once the method has been established, any data (such as selectivity measurements, or replicate measurements) collected during its development may be used as part of a validation study, where reasonable. Note also that substantive changes made to the procedures between method establishment and validation completion may affect the applicability of the data collected and therefore invalidate the validation.

C. Analytical test method development recommendations

When analytical testing is used to comply with a requirement in chapter IX of the FD&C Act or FDA's implementing regulations, submitters should use an analytical test method that has been validated (see section V of this guidance for reference). Analytical test methods should be performed under optimized conditions and within the limitations of the analytical test method. Analytical test methods should meet system suitability requirements (if applicable) to ensure that the complete testing system (including instruments, reagents, equipment, and laboratory personnel) is suitable for the intended purpose of analysis. Doing so will help ensure the data produced and submitted to FDA is accurate and reliable for FDA to use in making determinations regarding the protection of the public health.

Reference standards (e.g., calibration standards, internal standards, system suitability standards, and controls) should be used when executing system suitability testing, validation, and verification of analytical test methods. To have optimal method suitability, FDA recommends a certified tobacco reference standard be used when available.

System suitability acceptance criteria should be determined during the analytical test method development or optimization phase. Thereafter, analysts should determine that a method meets the system suitability criteria before accepting the results generated from an analytical test method. The frequency of system suitability evaluation should be determined by the analyst, with an understanding of the risk of discounting analytical data obtained while the system was not shown to be suitable. Typically, labs determine system suitability daily and have check standards interspersed during longer sample runs.

Laboratories should document activities and ensure that method and data quality criteria are met, including operational maintenance and performance verification of instruments; certificates of analysis for reagents and standards; and properly labeled laboratory solutions with established expiration dates. Solutions that have expired are not recommended to be used for any regulatory analytical testing, including validation and verification, as these solutions can add a level of uncertainty and lead to unreliable results.

Laboratories should use a laboratory reagent blank with every analytical test to detect potential contamination during the test material preparation and analysis process. The reagent blank or blanks are used to identify any cross contamination and interference prior to samples being run. The analyte being measured should be absent or below the limit of detection in the laboratory reagent blank for the particular method used. If the analyte being measured in the test material is detected in the laboratory reagent blank, this would skew the amount of unknown analyte in the test material, creating unreliable results. As such, it is recommended that the results of the analysis utilizing contaminated test material should not be reported and the cause(s) of the contamination should be properly diagnosed and corrected before the analysis is repeated on the laboratory reagent blank and test material.

D. Statistical analysis

Statistical analyses should be used to evaluate samples, analyze data, and both verify and validate performance characteristics. Statistical analyses of data generated during verification should be evaluated against the predetermined analytical test method validation performance criteria. All statistical procedures and parameters used in the analysis of the data should be based on sound statistical principles suitable for the intended evaluation.

V. ANALYTICAL MEASUREMENT PROCEDURE

A. General Recommendations

Analytical measurement procedure (AMP) validation experiments are conducted using solutions of the analyte or analytes in representative matrices (matrices that are similar to the composition of a sample that will be subject to the procedure). The analytical measurement procedure starts with a solution immediately following the final extraction/solubilization of the analyte from the sample matrix (e.g., extracts from a Cambridge filter pad, extract from tobacco filler, DNPH solution, etc.). Therefore, all of the filtration, centrifugations, liquid/liquid extractions, and dilutions should be considered as error sources in the evaluation of an analytical measurement procedure. For example, the extraction efficiency, grinding, and puffing protocols should not be included in this step, because they are components of the analytical sampling procedure validation. Replicates for AMP validation are prepared from separate aliquots of the collected final extraction/solubilization solution.

The validation parameters recommended to complete the validation of the AMP include: accuracy, precision (repeatability, intermediate precision, robustness), selectivity, linearity, range, and specificity (limit of quantification, and limit of detection).

B. AMP Specific Recommendations

Accuracy measurements for the AMP are collected using replicate standard solutions. The accuracy is based on replicate (seven or more is recommended, see section IV.A.1 above) standards or standard-spiked solutions. The accuracy is typically reported as the % recovered and is calculated from the mean value of the analyte as measured in the replicates and the known concentration of the analyte in the solution. For example, the data included in Table 2 above presents the % recovery of solutions having a known concentration. The % recovery is

calculated by dividing the mean measured concentration by the known concentration and multiplying the result by 100%. Generally speaking, increasing the number of replicates will result in a better estimation of any systematic error in the system and therefore is a better estimate of the true value. The number of replicates should always be enough to meet the predetermined criteria of the validation. The analyst should consider the effects of replicates prior to starting validation, and the rationale for choosing the replicates used should be documented.

Repeatability (precision) measurements for the AMP may be collected from the measurements used in the accuracy parameter above. The repeatability is reported as a % RSD and is calculated from the mean value and the standard deviation of the measurements. An example of this approach is demonstrated in the examples of total error probability above. Alternatively, in cases where an analyst wishes to minimize the number of replicates and the method is expected to be well within the expected performance and acceptance criteria, calibration curves collected on seven different days could be used for the accuracy, repeatability, linearity, and intermediate precision measurements. However, this approach does represent an increased risk of failing a total error probability calculation because of variability introduced by day-to-day variability.

Intermediate precision for the AMP is intended to probe the typical changes that may occur with an actively used procedure in a single laboratory. Therefore, recommended components of the intermediate precision experiment include changes caused by differences in instruments, analysts, and different days. In cases where there is only one instrument in a laboratory, it is possible to focus the intermediate precision experiment only on the analyst and day. Regardless of the combination of the variables, FDA recommends no fewer than six independent scenarios be used to complete the analysis. Each scenario should be separately collected from a different AMP sample manipulation. Each of these sets should start by utilizing the sample manipulations from the same working solution.

Examples statistically demonstrating some considerations for intermediate precision conditions: Assuming that a lab has two qualified analysts (A and B), two calibrated instruments (α and β) and several days (numbers 1-6) in which to complete this experiment. The lab might instruct:

- Analyst A to use instrument α on six days (Aα1, Aα2, Aα3, Aα4, Aα5, Aα6)
- Analyst A to use instruments α and β on three days (Aα1, Aβ1, Aα2, Aβ2, Aα3, Aβ3)
- Analysts A and B to use instrument α on three days (Aα1, Bα1, Aα2, Bα2, Aα3, Bα3)
- Analysts A and B to use instruments α and β on two days (Aα1, Bα1, Aβ1, Bβ1, Aα2, Bα2, Aβ2, Bβ2)
- Other variations may also be suitable

If the lab instructs the following, it would likely not provide an adequate number of independent scenarios:

- Analyst A to use instrument α for three days and Analyst B to use instrument β on same three days (Aα1, Aα2, Aα3) and (Bβ1, Bβ2, Bβ3). The two combinations are

independent but are not part of the same sample set and, therefore, would be two sets of three determinations, rather than one set of six.

- Analyst A to use instrument α, twice a day for three days. Multiple experiments by a single analyst typically do not provide insight into the longer-term variability that this experiment attempts to emulate.
- Analyst A and B to use instrument α for two days (Aα1, Bα1, Aα2, Bα2)

Intermediate precision is reported as a % RSD and is calculated from the mean value across all of the conditions tested and the standard deviation of the measurements in the same manner as the repeatability evaluation described above.

Robustness is typically evaluated using a set of experiments intended to identify the boundaries of acceptable instrument setting adjustments that can be made without causing a change in the AMP. In these experiments, the analyst systematically modifies instrument conditions and settings until the resulting data no longer meet the performance criteria or exhibit other issues (e.g., peak fronting, baseline anomalies, peak broadening, peak splitting, etc.). In many cases, changes to certain settings may have no detrimental effects on the AMP. In these cases, the validation report should note the conditions that were adjusted and the extent to which the adjustments did not affect or change the resulting data, with special attention given to the precision, accuracy, and selectivity. In cases where the adjustments do result in changes to the data, especially with regard to performance criteria, the condition or setting of each adjustment and the resulting changes should also be reported so that this information can be considered if the validated procedure is extended to other products or analytical applications.

Selectivity of an AMP may be accomplished through the evaluation of solutions containing the analyte and the potential interferences. Where an interference is detected, it would be useful to describe how the interference will be addressed. In cases of a chromatographic AMP, selectivity can be obtained by demonstrating acceptable chromatography through the analysis of resolution between a pair of peaks in the chromatogram (these two peaks are termed the critical pair). The critical pair may be the analyte and the nearest eluting interferent identified during method development. Typically, a solution containing the two compounds of the critical pair are prepared from standard reference materials or are a standard mixture. This solution is used as a system suitability solution and is often also used as a system check between sample injections. The critical pair may be two other peaks found to elute in close proximity to one another. If a critical pair does not include the analyte, data showing that the analyte peak is resolved should be provided. Where a multi-dimensional (multi-channel) detector is used (such as a photodiode array or tandem mass spectrometer) a demonstration that interfering compounds do not result in a determinate error should be provided.

The linearity and range of the AMP are typically evaluated at five or more concentrations over the linear dynamic range identified during method development. During the validation of the AMP, the solutions measured should be prepared from certified reference materials of the analyte in matrix matched solvents. In this case, the matched matrix to be used should be the solvent blank. The data obtained in the collection of the LOQ and repeatability values may be

used in the linearity and range determinations. The acceptability of the linearity and range will be evaluated by calculating the total error probability at each concentration using the repeatability values from Table 3 below. Unless otherwise indicated, the target concentration of the procedures will generally be assumed to be the median of the concentrations included in the linearity and range data sets. The slope, intercept, and degree of fit calculation (such as R^2) may be reported but are not generally used in determining the validity of the AMP because these parameters are only loosely correlated to measurements of error.[10]

The limit of quantification of the AMP is typically determined through serial dilutions of low concentrations of the analyte in solvent blank. As indicated above, the LOQ is directly related to the target concentration of the procedure. For example, if an AMP is intended to measure an analyte at 100 µg/mL, an acceptable LOQ will be different from one intended for a 100 ng/mL, as validation is a determination of the suitability of a method for the intended purpose. This is because the expected amount of error is greater at lower concentrations, which is reflected in the total error probability. See section VII for further discussion of setting acceptance criteria. Where the AMP is intended for multiple target concentrations, LOQs for each target concentration may be useful and are recommended.

The limit of detection is calculated using a solvent blank for the AMP validation. The calculated LOD represents the lowest amount of analyte that can be detected without typical matrix effects or sampling errors. This LOD represents what can be achieved by a typical laboratory and defines the limits of the measurement system and solution preparation processes.

VI. TOBACCO PRODUCT ANALYTICAL SAMPLING PROCEDURES

A. General Recommendations

The analytical sampling procedure (ASP) in most cases will be the source of the greatest variability in measurements that cannot be credited to the variability of the tobacco product itself. Validation of the ASP cannot occur without an associated AMP. However, any AMP that meets the acceptance criteria of the ASP validation should provide equivalent results to an ASP. The suitability criteria of the ASP differ slightly from that of the AMP, because of the increased number of error sources encountered during sampling. For example, an analytical sampling procedure intended to measure tar in cigarette smoke will contribute greater error to the measurement than the AMP and, therefore, the acceptable error of the analytical method will be defined by the ASP.

Evaluations of the ASP are based on the effects of both the sampling system and any interferences contributed by the tobacco product that have an effect on the analytical measurement. Therefore, a validation of the ASP should include comparative measurements of sample extracts/smoke from a certified reference product (e.g., 1R6F, CM8, CRP1, etc.) and of the tobacco product category that is the subject of the validation. The use of reference products allows comparison to known values, thereby reducing the number errors and better reflecting the reliability and suitability of the procedure. ASPs are generally developed to be applicable to all

[10] Ermer, J. and P. W. Nethercote (2014). Method Validation in Pharmaceutical Analysis. Singapore, Wiley-VCH.

of the tobacco products in the same category (e.g., smokeless tobacco, cigarettes, electronic nicotine delivery systems). However, the target levels of analytes can differ dramatically and differences in ingredients and product design may lead to substantial differences in the types and amounts of interferences collected by the ASP. Thus, each ASP should be verified or validated and included for each tobacco product analytical method in each regulatory submission for a tobacco product. The topic of verification is discussed at greater length in section IX.

The validation parameters recommended to complete the validation of the analytical sampling procedure generally include accuracy and precision (repeatability, intermediate precision, robustness), using the certified reference product if available.

B. ASP Specific Recommendations

Accuracy measurements for the ASP are made using a spike recovery approach. Often it is difficult to prepare a solution from a reference tobacco product that does not contain the analyte. Therefore, the accuracy value should be a relative content comparison. The reference product is smoked/aerosolized/extracted and the resulting solution (after extraction manipulation) serves as the blank sample solution. The blank sample solution is then spiked using a certified reference standard of the analyte at approximately the target concentration. The difference between the blank sample solution and the spike sample solution represents the accuracy calculated as a % recovery. The accuracy is based on replicate spiked solutions within the AMP linear dynamic range. Generally, seven replicates are sufficient to provide accuracy information.

Repeatability measurements for the ASP may be collected from the measurements used in the accuracy parameter above. The repeatability is reported as a % RSD and is calculated from the mean value and the standard deviation of the measurements.

Intermediate precision for the ASP is intended to probe the typical changes that may occur with the procedure in a single laboratory. Therefore, changes caused by differences in instruments, analysts, and different days are individual components to include in the intermediate precision experiment. Because ASP is intended to characterize the sources of error in the sampling procedure, each of the precision measures collected is specific to the sample collection. In cases where cigarette smoking is the sampling procedure, the instrument is the smoking machine. Multi-port smoking machines have the potential to add a variable systematic sampling error when there are differences in the ports. Therefore, when a multi-port smoking machine is used, it is appropriate to evaluate the systematic error present in the smoking machine used to collect HPHC data. Where this evaluation indicates a difference in systematic error it would be useful to describe how differences in the ports are addressed in a regulatory submission for a tobacco product. In cases where there is only one instrument in your laboratory, it is possible to focus the intermediate precision experiment only on the analyst and day. Regardless of the combination of the variables, we recommend that at least six independent scenarios be studied to complete the analysis (see the description and examples in the AMP section above for further details). As suggested in the example, each scenario might be collected from a different ASP sample manipulation (with corrections as appropriate), and separately collected.

Robustness is typically evaluated using a set of experiments that seeks to identify the boundaries of acceptable instrument (e.g., smoking machine) setting adjustments that can be made without causing a change in the ASP. In these experiments, the analyst systematically modifies instrument conditions and settings until the resulting data no longer meet the performance criteria or the result exhibits other significant issues. In many cases, changes to a certain setting have no detrimental effects on the ASP. In these cases, the validation report should note the conditions that were adjusted and the extent to which the adjustments yielded no adverse effects. In cases where the adjustments do result in changes to the data, especially with regard to performance criteria, the condition or setting of each adjustment and the resulting changes should also be reported. This portion of the validation is typically concluded as part of the method development. When available from that source, no additional data is needed, but the results should be reported from the method development information.

The limit of detection is calculated using a sample blank for the ASP validation. The calculated LOD represents the lowest amount of analyte that can be detected with representative matrix effects and sampling errors. This LOD represents the practical limits that a method can achieve for the product and method being validated.

VII. ANALYTICAL TEST METHOD VALIDATION ACCEPTANCE CRITERIA

A critical component of a validation experiment is the determination of the acceptance criteria necessary to demonstrate that the validation procedure is suitable for its intended purpose. This determination should be made prior to the beginning of the validation experiments and should be consistent across methods utilized for validation. The acceptance criteria selected for a validation need not be associated with the acceptance criteria of a tobacco product. For reference, the validation acceptance criteria applied in the evaluation of FDA-developed methods are based upon the work of William Horwitz[11] and are applied in a manner similar to that described in the FDA Food Program Validation guidance.[12]

A. Total Error Acceptance Ranges

Generally, the way to evaluate the acceptability of the measured error components for a given analytical method is the total error probability. The total error probability calculation is a statistical calculation that determines the probability that a measurement having the target concentration would fall within an acceptance criterion. The total error acceptance criterion should be defined prior to beginning the validation study and should be linked to concentration of the analyte in the sample matrix and criticality of the result. The lower the concentration, the wider the acceptance criteria should become. For example, nicotine in burley tobacco leaf is generally between 4 to 12 micrograms per gram of tobacco. However, an assay may measure as low as 0.1 microgram and as high as 30 micrograms and at either limit, it is scientifically recognized that accuracy will suffer. It is important to get the highest accuracy within the known range in which the analyte is likely to be seen. However, a generally accepted approach to

[11] Horwitz, W. (1997). "The variability of AOAC methods of analysis as used in analytical chemistry." J. Assoc. Off. Anal. Chem. **60**: 1355-1363.

[12] FDA. Foods Program (2019). Guidelines for the Validation of Chemical Methods in Food, Feed, Cosmetics, and Veterinary Products. https://www.fda.gov/media/81810/download.

estimate the acceptance criteria is provided by the Horwitz-Thompson equation and the repeatability.[13] As indicated in the prior sections, total error measurements for AMP validation (including accuracy, repeatability, intermediate precision, robustness, and LOQ) use the repeatability value %RSD to calculate the upper and lower values. Total error measurements for the ASP validation (including accuracy, repeatability, intermediate precision, robustness, and LOQ) use the reproducibility value %RSD to calculate the upper and lower values. The values provided in the table below are for illustrative purposes only and do not represent acceptable differences between tobacco products or acceptable levels of HPHCs or other constituents in a tobacco product for any reporting submission, record, or document under chapter IX of the FD&C Act or associated regulations.

Table 3: Concentration and variability criteria

Concentration of Analyte (%)	Concentration of analyte (ppm or ppb)	Concentration with w/w units (mg/g or µg/g)	Concentration Fraction	Reproducibility Value (% RSD)	Repeatability Value (% RSD)
10		1000 mg/g	0.1	2.8	1.4
1		10 mg/g	0.01	4	2
0.1	1000 ppm	1 mg/g	0.001	5.7	2.8
0.05	500 ppm	500 µg/g	0.0005	6	3
0.01	100 ppm	100 µg/g	0.0001	8	4
0.001	10 ppm	10 µg/g	0.00001	11	6
0.0001	1 ppm	1 µg/g	0.000001	16	8
0.00001	100 ppb	0.1 µg/g	0.0000001	22	11
<0.00001	< 100 ppb	< 0.1 µg/g	< 0.0000001	22	22

VIII. REFERENCE STANDARDS

A. General Reference Standards

All reference standards should be suitable for their intended use. All reference standards and prepared solutions should be properly labeled and stored with established expiration dates. Reference standards and solutions that have expired should not be used for any regulatory analytical testing, including validation and verification. All reference standards and prepared solutions should be accurate and reliable.

B. Tobacco Product Reference Standards

The tobacco product reference standard should be for the same product type as the test material. In addition, the tobacco product reference standard should be homogeneous and used concurrently with every test to verify accurate determination of the analyte(s). FDA recommends use of certified tobacco product reference standards.

[13] Massart, D. L., et al. (2005). "Benchmarking for analytical Methods: The Horwitz curve." Lc Gc Europe **18**(10): 528.

IX. ANALYTICAL TEST METHOD VERIFICATION OF TOBACCO PRODUCTS

Once a method has been validated, any subsequent substantial[14] change results in a "new" method that should be independently validated to ensure that the changes did not adversely affect the suitability of the method. The extent of the validation for the "new" method depends upon the extent of change.

Generally, validation should be conducted relative to a reference product similar to the product that is to be tested. The completed validation is applicable to the tobacco product evaluated in the validation process and can be extended to other products within the same category (e.g., cigarette, tobacco filler, etc.) through a verification process. For example, validation of an analytical method using a reference cigarette (such as CTRP 1R6F) does not mean that the validated procedure is suitable for the testing of other cigarette products. However, the validated method can be demonstrated to be applicable to another cigarette brand by completing a verification process. The verification process should demonstrate that the analyte of interest (e.g., nicotine or other HPHC) falls within the linear dynamic range of the validated method and that no interferences are present in the product under test that were not found in the reference product.

Verification is a shortened form of a validation procedure that is used to demonstrate that a previously validated method is suitable for use with another tobacco product in the same category or in some cases where different instruments are used (typically in cases where the instrument comes from a different vendor). The verification parameters recommended include: accuracy and selectivity of the sample tobacco product subject to the applied method. Additional verification parameters may be added at the analyst's discretion. Verification should be completed for each analyte intended to be measured with the analytical method to demonstrate the assay performs as expected and provides consistent accurate results. Verification is similar to the ASP validation approach. As with validation, performance criteria as strict as those typically used for purposes of validation should be determined prior to conducting the verification experiments.

Accuracy measurements for verification are made using a spike recovery approach. Often it is difficult to prepare a solution from a reference tobacco product that does not contain the analyte. Therefore, the accuracy value should be a relative content comparison. The reference product is smoked/extracted and the resulting solution (after extraction manipulation) serves as the blank sample solution. The blank sample solution is then spiked using a certified reference standard of the analyte at approximately the target concentration. The difference between the blank sample solution and the spike sample solution represents the accuracy calculated as a % recovery. The accuracy is based on replicate spiked solutions. FDA recommends using seven or more replicates at a variety of concentrations within the AMP linear dynamic range.

[14] Examples of the types of changes that could be considered "substantial" are described at length for chromatographic procedures in the current *USP/NF* under General Chapter Chromatography <621>, section "system suitability."

Selectivity for verification may be measured by evaluating solutions containing analyte and potential interferences. Where an interference is detected, it would be useful to describe how the interference will be addressed. In cases of a chromatographic AMP, selectivity can be obtained through the demonstration of acceptable chromatography through the analysis of resolution between a pair of peaks in the chromatogram (these two peaks are termed the critical pair). The critical pair may be the analyte and the nearest eluting interferent identified during method development. Typically, a solution containing the two compounds of the critical pair are prepared from standard reference materials or are a standard mixture. This solution is used as a system suitability solution and is often used as a system check between sample injections. The critical pair may be two other peaks found to elute in close proximity to one another. If a critical pair does not include the analyte, data showing that the analyte peak is resolved should be documented. Where a multi-dimensional (multi-channel) detector is used (such as a photodiode array or precursor or product ion mass spectrum of the analyte peak), the analyst should be able to demonstrate that interfering compounds do not result in a determinate error.