Guidelines for Validation of Quantitative Gluten Methods, with Specific Examples for ELISA Assays

1 Scope

The purpose of this document is to provide comprehensive technical guidelines for method developers conducting validation studies for quantitative gluten methods, for example methods submitted for AOAC INTERNATIONAL (AOAC) *Performance Tested Methods*℠ (PTM) status and/or for AOAC *Official Methods of Analysis*℠ (OMA) status. This document is not intended to describe requirements for laboratories using commercial methods for gluten analysis, though for these laboratories it would assist their understanding of the consensus-based approach, the terminology used, and what information they can expect to receive from method developers.

The requirements for method developer single-laboratory validation (SLV) studies, independent validation studies, and collaborative validation studies are described. Specific examples are provided for Enzyme-Linked Immunosorbent Assay (ELISA) methods.

For AOAC PTM and OMA validations, a study protocol should be reviewed prior to commencement of the study.

2 Applicability

These guidelines are intended to be applicable to the validation of candidate quantitative gluten methods, whether proprietary or non-proprietary, including those that may be submitted to AOAC for OMA status or PTM certification. Unforeseen circumstances may necessitate divergence from these guidelines in certain cases, and these must be reviewed by AOAC or another appropriate agency (other than the method developer). The AOAC PTM Program requires a method developer SLV, and an independent laboratory study. The AOAC OMA Program requires an SLV (also known as the pre-collaborative study) and a collaborative study to achieve Final Action status. A harmonized PTM-OMA Program can be followed in which PTM certification is sought and, if successful, serves as the SLV phase of the OMA Program.

3 Terms and Definitions

Where appropriate, definitions have been taken from international standards and the source is cited. Sources of definitions and other references are included in the Reference list.

(a) *Analyte.*—Chemical entity or entities measured by the measurement system, which may be a marker (e.g., a specific gluten peptide or protein) or a surrogate (e.g., another protein from wheat, rye, barley or oats that correlates with the presence of gluten).

See also "Measurand" definition. *See* De Bievre (1) for a detailed discussion of the difference between "analyte" and "measurand".

(b) *Bias.*—Difference between the expectation of the test results and an accepted reference value. Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias.

(c) *Calibrant.*—A material used for calibration of a measurement procedure.

(d) *Candidate method.*—The method submitted for validation.

(e) *Candidate method result.*—The final results of the quantitative analysis for the candidate method.

(f) *Collaborator.*—An intended user who participates in the collaborative study.

(g) *Cross-reactivity.*—A measurable response, above the LOQ of the method, to a material other than the target analyte.

(h) *Cross-reactivity study.*—The examination of matrices that do not contain claimed analyte, which are potentially cross-reactive, to determine that they do not produce a measurable response above the claimed LOQ of the method.

(i) *Enzyme-linked immunosorbent assay (ELISA).*—An analytical procedure characterized by the recognition and binding of specific antigens by antibodies and signal generation by an enzyme-substrate reaction.

(j) *Gluten.*—A protein fraction from wheat, rye, barley, oats or their crossbred varieties and derivatives thereof, to which some persons are intolerant, and that is insoluble in water and 0.5M NaCl (2), Throughout this document, the word 'wheat' refers to all Triticum species and their crossbreeds, such as triticale, durum wheat, spelt and Khorasan wheat, and their hybrids and crossbred varieties such as Triticale. [Per Codex Standard 119-1979, "oats can be tolerated by most but not all people who are intolerant to gluten. Therefore, the allowance of oats that are not contaminated with wheat, rye or barley in food [...] may be determined at the national level (2)].

(k) *Incurred test material.*—A material prepared from a food matrix into which a gluten source (e.g., flour) has been incorporated prior to subjecting the matrix to a given food processing operation.

(l) *Independent testing site.*—A testing site not owned, operated or controlled by the same entity as the method developer.

(m) *Interference study.*—The examination of matrices expected to be tested with the method, to demonstrate that they do not interfere with detection of the analyte.

(n) *Intermediate precision.*—Precision under intermediate conditions (ISO 3534-2; 3). For the purposes of this document, the subscript notation "i" will be used to indicate terms and estimators associated with intermediate precision. Estimation methods can be found in Section **4.6**

(o) *Intermediate precision conditions.*—Conditions where test results or measurement results are obtained with the same method, on identical test/measurement items in the same test or measurement facility, under some different operating condition, which may include, but are not limited to: time, calibration, operator, reagent lots and equipment.

Specific criteria for intermediate precision conditions are given in Section **4.4**

(p) *Limit of detection (LOD).*—The lowest concentration or mass of analyte in a test material that can be distinguished from a true blank test material at a specified probability level (ISO 5725-1:2023; 4). *See* further details on how to determine LOD in Section **6.5**.

(q) *Limit of quantification (LOQ).*—The lowest level of analyte in a test portion that can be reasonably quantified at a specified level of precision (ISO 5725-1:2023; 4). *See* further details on how to determine LOQ in Section **6.6**

(r) LOQ_{RSD}.—A limit of quantification with a specified intermediate precision relative standard deviation, expressed as a percentage. For example, an LOQ_{10} from a single laboratory validation would be the lowest concentration where the $RSD_i = 10\%$, and the LOQ_{10} from a collaborative study would be the lowest concentration where the $RSD_R = 10\%$.

(x) *Matrix.*—Totality of components of a material system except the analyte (ISO 17511; 5). For example, the food, beverage, or environmental surface material to be included in the validation as per the intended use of the method.

(y) *Measurand.*—The quantity intended to be measured (the specification of the measurand should be sufficiently detailed to avoid any ambiguity). *See* also "analyte" definition.

(z) *Measurement interference.*—A cause of significant bias in the measured analyte concentration due to the effect of another component or property of the sample which may result from non-specificity of the detection system, suppression of an indicator reaction, or inhibition of the analyte. (CLSI guideline EP07-A2; 6) An interference can be endogenous, present in the sample, or exogenous, introduced into the sample during the measurement process.

(aa) *Measurement range.*—The concentration range over which the target analyte can be reliably quantified/detected.

(bb) *Precision.*—The closeness of agreement between independent test results under stipulated conditions. (ISO 5725-1; 4).

(cc) *Qualitative method.*—Method of analysis whose response is either the presence or absence of the analyte.

(dd) *Quantitative method.*—Method of analysis whose result is the amount (mass or concentration) of the analyte.

(ee) *Recovery.*—The fraction or percentage of analyte that is recovered when the test portion is analyzed using the entire method.

(ff) *Reference material.*—Material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process (*see* NIST SRM Definitions https://www.nist.gov/srm/srm-definitions).

(gg) *Repeatability.*—Precision under repeatability conditions. (ISO 5725-1; 4).

(hh) *Repeatability conditions.*—Conditions where independent test results are obtained with the same method on equivalent test items in the same laboratory by the same operator using the same equipment within short intervals of time.

(ii) *Reproducibility.*—Precision under reproducibility conditions (ISO 5725-1; 4).

(jj) Reproducibility conditions.—Conditions where independent test results are obtained with the same methods on equivalent test items in different laboratories with different operators using separate instruments.

(kk) *Robustness.*—Measure of the capacity of an analytical procedure to remain unaffected by small variations in method parameters; provides an indication of the method's reliability during normal usage.

(ll)*Selectivity.*—The degree to which the method can quantify the target analyte in the presence of other analytes, matrices, or other potentially interfering materials. Includes:

(1) Breadth.—The ability of the method to detect gluten from multiple grain sources.

(2) Cross-reactivity.—See definition of cross-reactivity above.

(3) Measurement interference.—A cause of significant bias in the measured analyte concentration due to the effect of another component or property of the sample which may result from non-specificity of the detection system, suppression of an indicator reaction, or inhibition

of the analyte (CLSI_EP07-A2; 6). An interference can be endogenous, present in the test material, or exogenous, introduced into the test material during the measurement process.

(mm) *Spiked test material.*—A food matrix into which gluten has been incorporated after all relevant food processing operations have been completed (*see* **Annex A** for details).

(nn) *Test material.*—A material used for method validation that either contains a gluten source present at a given concentration in the context of a food or environmental matrix or is a blank matrix free of gluten.

(oo) *Test portion.*—Portion of the test sample as prepared for testing or analysis, where the whole quantity is used for analyte extraction at one time. (ISO 16577:2022; 7)

4 Method Developer Validation Study

Quantitative methods are those whose result is the amount (mass or concentration) of the analyte. This guidance has been developed for use with candidate methods that are designed to quantify gluten. If a candidate method's intended use is not covered by this document or existing *Standard Method Performance Requirements* (SMPRs^{SM}), the standing AOAC expert review panel (ERP) for gluten, or other qualified agency, may determine the appropriate cross-reactivity/interference panels, and performance requirements.

Method developers may prepare study test materials in-house for the SLV (method developer study), but all test materials and test portions must be blind-coded and randomized. Analyses conducted by the method developer must be performed by an independent analyst without prior knowledge of the test materials undergoing analysis. Ideally, all test materials for the independent laboratory and collaborative studies should be prepared by an external entity independent from the method developer. At least one incurred test material for the independent laboratory and collaborative studies must be prepared by an external entity independent from the method developer. In situations where an independent entity is unavailable to prepare all of the test materials for the independent laboratory and collaborative studies, or their use is impractical for all test materials, method developers may produce and distribute test materials as long as detailed information is provided on procedures used to prevent bias (preparation, coding, etc.), and justification is provided for failing to use an independent entity to prepare all of the test materials.

4.1 Scope

A SLV study (also referred to as a Method Developer Study), is intended to evaluate the performance of a candidate method in the following areas: (*1*) calibration fit, (*2*) selectivity, (*3*) precision (repeatability and intermediate precision), (*4*) LOD/LOQ, (*5*) recovery, and (*6*) robustness. These studies are generally conducted within a method developer laboratory.

Gluten has multiple potential sources – wheat, rye, barley, oats and their hybrids and crossbreeds – and multiple regulatory levels. Developers must determine which of these sources and levels their method is intended to detect, and perform matrix studies for each claimed gluten source.

4.2 Calibration Fit Study

Analyze calibration standards as they are included in the test kit, or prepared as described in the test method. Analyze at least four replicates of each concentration defined for the calibration curve. Fit the calibration curve using the regression model described in the method instructions and/or kit insert, plotting each individual data point and not averaging. Full descriptions must be provided with respect to performing the calibration function calculations, including any

transformations conducted and the regression model used. Full calibration curve plots and calibration functions must be shown.

From the calibration curve function, determine the calculated concentrations for each of the standards. Calculate the residuals for each concentration standard for each replicate. Residuals are the difference between the observed value and the predicted value for each dependent variable in the calibration curve. (Residual = observed value - predicted value.) Residuals should be calculated from the instrument response. For most quantitative gluten methods, instrument response would be optical density (absorbance) values.

Plot the residuals versus concentration. Residuals should have random distributions and be centered on zero. If a non-random pattern is observed, the calibration function or measurement range may not be appropriate. Residuals should generally also be <15% of the measured response, and up to 20% at the lowest non-zero calibration standard.

4.3 Selectivity Study

The selectivity study is intended to provide information on potential sources of cross-reactivity and interference. The information related to cross-reactivity and interference should be reported in the validation report or in the package insert from the method developer.

(a) *Breadth*.—This section of the validation is intended to provide information to end users on the method's performance with less common varieties of gluten-containing grains, such as einkorn, spelt and emmer.

The materials identified in **Annex A**, Table A1, should be tested at three times the limit of quantitation (LOQ) of the method (as long as that is equal to or below 20 mg/kg, otherwise test at 20 mg/kg) in a rice flour matrix. Test six test portions per test material.

The absorbance or optical density (OD) values for all test portions and standards must be reported. The mean gluten concentration for each gluten source must be reported. Mean concentrations below the LOQ should be reported as below the limit of quantitation (BLQ). Percent recovery should be calculated and reported for the mean concentration from each gluten source. If any analysis is repeated, all datasets must be reported and a justification given for all repeat analysis.

For methods claiming wheat, only common wheat (*Triticum aestivum*) should be used in all other studies described in this guidance.

As the Breadth study is purely informational, there are no acceptance criteria, but method developer should point out any of the gluten-containing grains that demonstrate recoveries below 50%, in the method instructions.

(b) *Cross-reactivity*.—The matrices identified in **Annex A**, Table A2, at full, undiluted concentration (with some exceptions as noted), will be prepared and analyzed with the candidate method as it is designed for testing food products. One test portion of each blank food material should be analyzed according to the entire method protocol.

In the event that an unclaimed matrix tests above the method LOQ or lowest non-zero standard, it or another example of the same matrix may be retested in six test portions, to rule out cross-reactivity. If the result persists, the extract must be diluted and rerun to characterize the extent of the cross-reactivity, and the test material may also be evaluated with an alternative method (PCR, Western blot, mass spectrometry, alternate ELISA, etc.) to verify whether the signal is the result of cross-reactivity or a true positive due to cross-contact.

The absorbance or OD values for all test portions and standards must be reported. The extrapolated concentration for all test portions that had an absorbance or OD above the limit of quantitation of the method must be reported. If any analysis is repeated, all datasets must be reported and a justification given for all repeat analysis.

Any cross-reactive matrix must be reported to end user as part of the method instructions.

(c) *Interference.*—The matrices identified in **Annex A**, Table A2 will be spiked with gluten from each claimed gluten source at three times the LOQ of the method (as long as that is equal to or below 20 mg/kg, otherwise test at 20 mg/kg). Test material preparation is described in **Annex B**. One test portion of each spiked test material will be analyzed with the candidate method as it is designed for testing food products.

If a result is obtained that is above the measurement range of the method, the extract must be diluted and re-analyzed.

The absorbance or OD values for all test portion extracts and standards must be reported. The concentration for all test portions that had an absorbance or OD above the LOQ of the method must be reported. If any analysis is repeated, all datasets must be reported and a justification given for all repeat analysis. The percent recovery should be calculated and reported for each tested food.

Spiked test materials must render a result above the LOQ. In the event that the single test portion replicate tests below the LOQ, that food matrix may be retested in six additional test portions, with no results below the LOQ allowed, to rule out interference.

Findings that certain matrices interfere with gluten detection should be investigated further, using additional similar matrices, to determine the full scope of interference. Any interfering matrices must be reported in the method instructions.

4.4 Matrix Study

The matrix study is intended to provide data on precision (repeatability and intermediate precision), limit of detection (LOD)/LOQ, and recovery in a controlled laboratory setting for all gluten sources, matrices and surfaces claimed in the method's intended use statement.

A matrix study must be performed in each claimed matrix. In order to ensure that each claimed gluten source is represented, the gluten sources must be rotated across the claimed matrices as shown in Tables 1 or 2. The single matrix in which all gluten sources are tested, listed in Tables 1 and 2 as Matrix A, should be the most highly processed matrix used in the validation study.

Table 1. Rotation of gluten sources across claimed matrices for methods claiming to detect wheat, rye, and barley. The rotation of single gluten sources would continue for six matrices and greater.

	Number of matrices claimed				
	1	2	3	4	5
Matrix A	Wheat Barley Rye	Wheat Barley Rye	Wheat Barley Rye	Wheat Barley Rye	Wheat Barley Rye
Matrix B		Wheat	Wheat	Wheat	Wheat
Matrix C			Barley	Barley	Barley
Matrix D				Rye	Rye
Matrix E					Wheat

Table 2. Rotation of gluten sources across claimed matrices for methods claiming to detect wheat, rye, barley, and oats. The rotation of single gluten sources would continue for six matrices and greater

	Number of matrices claimed				
	1	2	3	4	5
Matrix A	Wheat Barley Rye Oats	Wheat Barley Rye Oats	Wheat Barley Rye Oats	Wheat Barley Rye Oats	Wheat Barley Rye Oats
Matrix B		Wheat	Wheat	Wheat	Wheat
Matrix C			Barley	Barley	Barley
Matrix D				Rye	Rye
Matrix E					Oats

Alternatively, a matrix study for a matrix category may be performed by testing each claimed gluten source, per the rotation shown in Tables 1 or 2, in at least five examples from the category, equally distributed across each available type of processing (**Annex C**). Test materials under each type of processing must be incurred. As an example, a method wishing to make a claim for the "Cereals (Not Fermented, Hydrolyzed or Fractionated)" category would need to test one matrix from each of the five provided processing categories, and in each instance, gluten would need to be added to the matrix prior to the described processing step. If a method developer was unable to access suitable equipment for preparing incurred test materials in the Pressure/Extruded type of processing, but was able to make incurred test materials for all other types of processing, they could not claim the "Cereals (Not Fermented, Hydrolyzed or Fractionated)" category. However, they could make a limited claim for "Raw, Processed, Baked, Fried and Dehydrated Cereals". Method developers with the ability to produce fermented, hydrolyzed or fractionated matrix test materials that were incurred with gluten prior to these processes may make individual claims based on the fermentation organism, hydrolyzing agent or fractionation process. Example claims would be "Soy Tempeh fermented with *Rhizopus oligosporus*", "Modified corn starch hydrolyzed with sodium hydroxide", or "Wheat starch fractioned with water".

Incurred test materials are required for evaluation of precision, LOD/LOQ, and recovery. *See* **Annex B** for description of best practices for incurred test material preparation.

At least four concentrations per matrix/gluten source combination, including a zero/blank, must be included in the study. The "Low" concentration should be less than or equal to two times the stated LOQ of the method, provided this is less than or equal to 20 mg/kg (if not, then the "Low"

concentration should be 20 mg/kg). Other concentrations should span the calibration range, e.g., at the middle and upper end of the calibration curve.

Individual studies may be designed for each performance parameter (repeatability, intermediate precision, LOD/LOQ, and recovery). Intermediate precision study designs must include multiple test portions, at least two test kit lots, and day/operator as a single confounded factor.

Alternatively, a single, statistically valid study may be designed and utilized to provide estimates of precision (repeatability and intermediate precision), LOD/LOQ, recovery, and lot-to-lot variability – *see* Figures 1–4 for examples of acceptable study designs, but other designs may also be able to give satisfactory data. Designs 1b and 2b (Figures 2 and 4) will provide sufficient data for all parameters in the *Matrix Study and the Product Consistency and Stability Study* (**5.2**), if conducted on a sufficient number of test materials. At least four concentrations per matrix/gluten source combination, including a zero/blank, must be included in these studies as well.

For methods that require the measurement of multiple replicate ELISA wells for each test portion, use Designs 2a or 2b (*see* Figures 3 and 4), or other designs that include replicate wells per test portion. For methods that only require the measurement of one ELISA well for each test portion, any of the four study designs may be used.

In order for the nested designs to be capable of estimating repeatability, at least two test portions must be analyzed under repeatability conditions (i.e., conducted on the same day, by the same operator, with the same calibration and equipment). Under these conditions, the nested designs can estimate both intermediate precision and repeatability because repeatability is a variance component within intermediate precision, as expressed in the following equation, where s_I^2 is the intermediate precision variance, s_{lot}^2 is the variance contributed by test kit lot, $s_{d/op}^2$ is the variance from the confounded factor of day and operator, and s_r^2 is the repeatability variance:

$$s\_I\text{^}2 = s\_lot\text{^}2 + s\_(d/op)\text{^}2 + s\_r\text{^}2$$

Repeatability estimates are required at four concentrations for each claimed matrix: blank, low, medium, and high levels, according to the claimed method quantification range.

As intermediate precision estimates are used for the calculation of LOD and LOQ, estimates are required for all matrices, with at least three concentration levels per matrix: blank, low, and medium.

Table 3. Required test materials for quantitative study designs

Parameter	Number of matrices and concentrations
Repeatability	All matrices, 4 concentrations (blank, low, medium, and high) for each matrix
Intermediate Precision	All matrices, at least 3 concentrations (blank, low, medium) for each matrix
LOD/LOQ	All matrices, at least 3 concentrations (blank, low, medium) for each matrix
Recovery	All matrices, three non-blank concentrations (i.e., low, medium, and high)

Test kit lot variance (lot-to-lot consistency) must be evaluated for at least one matrix using three test kit lots. This can be included in the estimation of intermediate precision (Designs 1b and 2b, Figures 2 and 4) or may be conducted separately (*see Robustness Study*).

Design 1a (Figure 1) can be used to estimate (*1*) intermediate precision (which includes repeatability, test kit lot variance (with 1 degree of freedom, df), and day/operator confounded variance) and (*2*) repeatability.

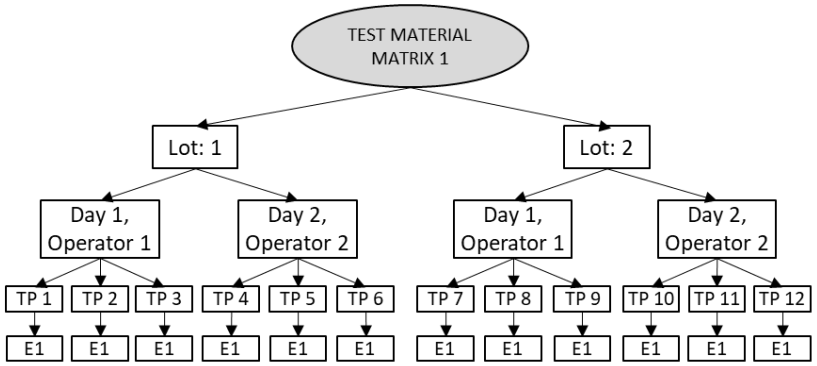


Figure 1. Design 1a. Lot: test kit lot, TP: test portion, E: ELISA measurement. Design 1a can be used to estimate intermediate precision and repeatability.

Two test kit lots are used to analyze each test material. Two operators conduct analysis on 2 days for each test kit lot. For each day and lot, the assigned operator conducts extraction and analysis of two test portions of the test material, with one ELISA measurement performed per test portion.

Design 1b (Figure 2) can be used to estimate (*1*) intermediate precision (which includes repeatability, test kit lot variance (with 2 df), and day/operator confounded variance) (*2*) repeatability, and (*3*) lot-to-lot product consistency.

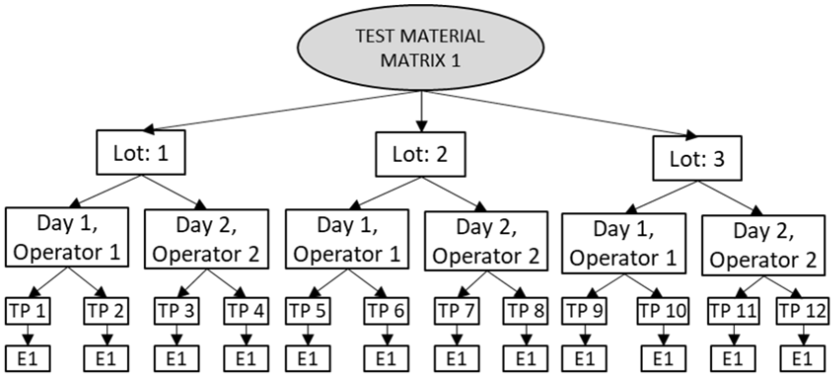


Figure 2. Design 1b. Lot: test kit lot, TP: test portion, E: ELISA measurement. Design 1b can be used to estimate intermediate precision, repeatability, and lot-to-lot product consistency.

Three test kit lots are used to analyze each test material. Two operators conduct analysis on 2 days for each test kit lot. For each day and lot, the assigned operator conducts extraction and analysis of two test portions of the test material, with one ELISA measurement performed per test portion.

Design 2a (Figure 3) can be used to estimate (*1*) intermediate precision (which includes repeatability, test kit lot variance (with 1 df), day/operator confounded variance, and ELISA variance), (*2*) repeatability (which includes test portion and ELISA variance), and (*3*) ELISA variance.

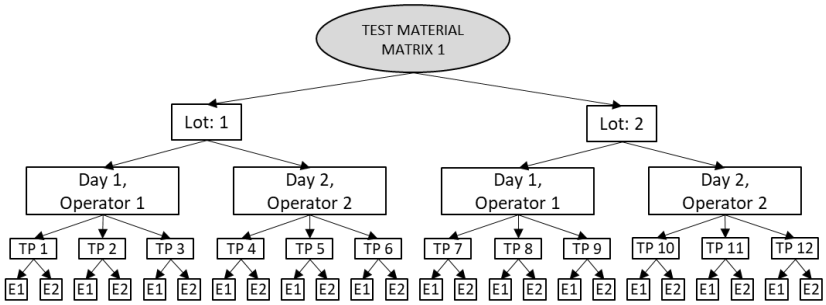


Figure 3. Design 2a. Lot: test kit lot, TP: test portion, E: ELISA measurement. Design 2a can be used to estimate intermediate precision, repeatability, and ELISA variance.

In this instance the repeatability variance can be further split into test portion variance and ELISA variance as shown in the equation below, where s_r^2 is repeatability variance, s_{TP}^2 is the variance attributed to test portion, s_{ELISA}^2 is the variance attributed to ELISA measurement variance:

$$s_r^2 = s_{TP}^2 + s_{ELISA}^2$$

Two test kit lots are used to analyze each test material. Two operators conduct analysis on 2 days for each test kit lot. For each day and lot, the assigned operator conducts extraction and analysis of two test portions of test material, with two ELISA measurements performed per test portion.

Design 2b Figure 4 can be used to estimate (*1*) intermediate precision (which includes repeatability, test kit lot variance (with 2 df), day/operator confounded variance, and ELISA variance), *(2)* repeatability (which includes test portion variance and ELISA variance), (*3*) ELISA variance, and (*4*) lot-to-lot product consistency.

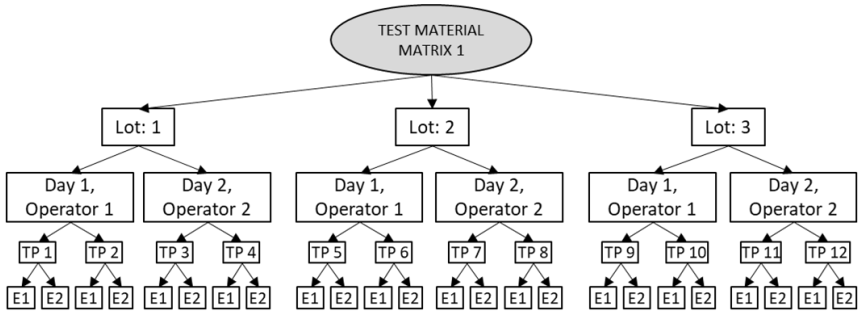


Figure 4. Design 2b. Lot: test kit lot, TP: test portion, E: ELISA measurement. Design 2b can be used to estimate intermediate precision, repeatability, ELISA variance, and lot-to-lot product consistency.

Three test kit lots are used to analyze each test material. Two operators conduct analysis on 2 days for each test kit lot. For each day and lot, the assigned operator conducts extraction and analysis of two test portions of test material, with two ELISA measurements performed per test portion.

If repeatability is conducted separately (Figure 5) at least six test portions of each test material should be analyzed according to the entire method as written. Analysis should be conducted by one analyst on 1 day, using one test kit lot and the same equipment (n = 6 per test material).

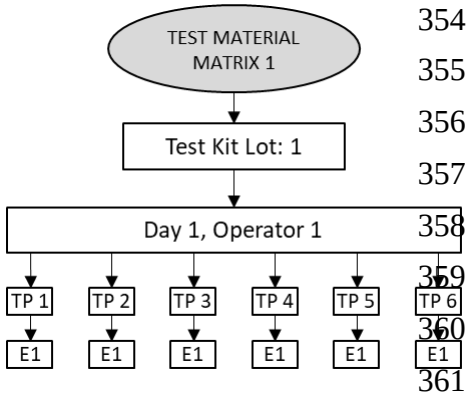


Figure 5. Repeatability only design.

(a) *LOD/LOQ estimation.*—In SLV studies for gluten immunoassay methods, the LOD and LOQ will be estimated using intermediate precision data.

Data collected from analysis of incurred test materials for all matrices will be used to model the relationship between analyte concentration and intermediate precision (*see* **Annex D**). Data used must meet other method performance criteria (e.g., recovery).

(b) *Recovery assessment.*—Data collected for the purposes of precision evaluation may also be used for the recovery assessment.

If conducted separately from the precision assessment, evaluate each incurred matrix with six independent analyses (test portions) per concentration level at a minimum of three non-blank concentration levels covering the analytical range.

4.5 Data Analysis and Reporting for Matrix Studies

(a) *Nested designs: repeatability and intermediate precision.*—Data generated from nested designs, such as those as described above, should be analyzed by an ANOVA capable of providing estimates of intermediate precision and repeatability. **Annex D** contains full instructions, R code, and example datasets for the study designs described in this guidance.

(b) *Repeatability only.*—In a situation where a study design for estimating repeatability alone is selected, the mean, standard deviation, and relative standard deviation should be calculated for each test material (i.e., each matrix-concentration combination). Formulas for standard deviation and relative standard deviation, as defined in OMA Appendix F (8), are as follows:

$$\text{Standard deviation } (s_r): s_r = [\Sigma(x_i - \bar{x})^2/(n-1)]^{0.5}$$

$$\text{Relative standard deviation (RSD): } RSD_r = s_r \times 100/\bar{x}$$

The study report must include the standard deviation and RSD values for each test material, and all repeatability estimates must meet requirements set forth in the relevant SMPR or established by the ERP or other review panel. In the absence of an SMPR and ERP, acceptable RSD_r values for gluten immunoassays are generally ≤20% within the claimed measurement range of the assay.

(c) *LOD, LOQ*.—LOD will be estimated using a hypothesis test approach, with $\alpha = \beta = 0.05$. The relationship between observed concentration and intermediate precision standard deviation must be taken into account in the estimation of LOD (also referred to as a precision profile estimation method for LOD). Full instructions for the calculations to estimate LOD are in **Annex D**.

LOQ estimation will be based on the relationship between concentration and intermediate precision standard deviation. Full instructions for the calculations to estimate LOQ are in **Annex D.**

LOD and LOQ can be estimated per gluten source and matrix, or as pooled values across all gluten sources and matrices if variances are homogeneous.

(d) *Recovery*.—

$$\text{Percent Recovery} = (\text{Experimental concentration})/(\text{Expected concentration}) \times 100$$

The expected concentration for each test material should be calculated from the incurred concentration, accounting for any mass changes during processing operations (e.g., moisture loss during baking).

For each claimed matrix and gluten source, plot the observed concentration vs. expected concentration for all levels, and perform a linear regression to determine the slope and confidence interval of the slope. Also calculate and report the recovery and confidence interval at each concentration, by taking the mean of the test portion values and calculating the recovery.

4.6 Acceptance Criteria for Matrix Studies

Each claimed gluten source (wheat, rye, barley and/or oats) in each matrix (or pooled across matrices if all matrices show equivalent recoveries) should all produce recovery values (determined as the mean value by weighted linear regression, with the associated confidence intervals) that comply with the relevant method performance requirements (e.g., AOAC SMPR). In the absence of an applicable SMPR, an ERP will evaluate the study data according to their expert opinions. With respect to recovery, while ideal values are from 80–120%, for single-gluten-source validations values of 50–150% can be acceptable [Abbott et al. (9)]. For multiple gluten source validations (e.g., wheat, rye and barley), values of 50–200% can be acceptable at the discretion of the ERP (AOAC SMPR 2017.021; 10). In the event that the confidence interval of the recovery mean as determined by weighted linear regression does not fall within the specified recovery range, the test material may be retested in additional test portions, and a new confidence interval calculated, to qualify as a gluten source quantified by the method. All data must be reported, included any testing done on different grain sources and varieties, and retests must be explained. Any gluten sources or matrices that do not meet these criteria cannot be claimed, and must be reported in the method instructions.

All parameter point estimates must meet any applicable requirements for confidence intervals established by the AOAC Statistics Committee or other relevant guidance.

If an applicable SMPR is available, the SLV study data must meet the corresponding criteria.

(a) *LOQ*.—The RSD_i at the LOQ must be less than or equal to the RSD_i in the relevant SMPR (or the RSD_R if an RSD_i is not listed). If there is no SMPR available for, RSD_i at the LOQ must be $\leq 30\%$.

If a method developer has an LOQ claimed as part of the method design (e.g., the lowest non-zero calibrant), the estimated LOQ from the SLV (which meets the SMPR requirements for maximum RSD_i) must be less than or equal to the claimed LOQ of the kit, within statistical

tolerances. If the estimated LOQ from the SLV is greater than the claimed LOQ of the kit, the method developer must revise the LOQ claimed in the test kit insert and validation reports to meet the precision requirements for LOQ.

In the validation reports and test kit inserts, the method developers must indicate the actual RSD_i value estimated for the LOQ of the kit as part of the LOQ information. For example:

LOQ_{15}, for a method where the existing LOQ claimed by the kit had an estimated RSD_i of 15% in the SLV

LOQ_{30}, for a method where the LOQ was set based on the SLV outcome and a maximum RSD_i of 30%. Acceptance criteria for the maximum RSD also includes meeting requirements for confidence intervals, as established by the AOAC Statistics Committee.

The LOQ estimate must be greater than or equal to the LOD estimate. If the LOQ estimate is lower than the LOD estimate, the LOQ should be reported as the same concentration as the LOD.

4.7 Robustness Study

The method developer, in conjunction with the AOAC or other independent validation manager, is expected to make a good faith effort to determine which, and to what magnitude, parameters are most likely to vary in the hands of an end user.

Analysis should be conducted on a minimum of one claimed matrix type, using one claimed gluten source.

Spiked matrices are acceptable for test kit lot-to-lot stability analysis and robustness analysis (except when varying extraction conditions). *See* **Annex B** for description of best practices for spiked matrix preparation.

Incurred matrices may also be used for the robustness study, and should be used if extraction conditions are varied. If sufficient quantities of incurred matrices have been prepared for the matrix study, these test materials may also be used for the robustness studies (i.e., separate incurred matrices are not required).

The robustness of the method should be investigated by performing experiments in which specific parameters are changed to determine the impact on the experimental result. In particular, the effect of deviations in incubation times, reagent volumes, extraction conditions (time and temperature) should be investigated. Each parameter should be varied both up and down by at least 20%. These parameters should be tested in a factorial or Plackett-Burman design, as described in **Annex D.**

Five test portions should be tested for a test material at three times the LOQ (as long as that is equal to or below 20 mg/kg, otherwise test at 20 mg/kg), and two test portions should be tested of a blank test material, for each treatment condition.

Data should be analyzed as described in **Annex D**, or by other appropriate ANOVA, multi-factor regression or generalized linear model software. If any of the experimental conditions evaluated significantly affect the results, this should be reported in the kit insert information as an instruction to end users to take special care not to vary that factor.

(a) *Product stability and consistency.*—If the test method is sold as a kit or device prepared in lots or batches, a product consistency and stability study must be performed to ensure that the performance of the product is consistent from lot-to-lot and over time under normal storage conditions for the shelf life of the product. Lot-to-lot consistency and product stability can be

measured in the same set of experiments. As specified in Section **4.4**, lot-to-lot stability and consistency can also be assessed in the context of nested designs for intermediate precision estimation that utilize at least three lots of test kits. Alternatively, method developers may provide internal lot-to-lot and stability data for review, as long as the volume of data meets or exceeds the data requested in the product stability and consistency studies described here.

The shelf life should include the stability of all the reagents provided with the test kit, ideally through real-time testing of reagents under normal storage conditions. Accelerated stability testing at higher than normal storage temperatures can also be used to estimate stability. An expiration date for each test kit should be clearly indicated, along with appropriate conditions for storage before use.

A minimum of three separate product lots must be evaluated. The product lots should span the shelf life of the kit. For example, if the kit shelf life is 12 months, an approximately 12-month-old kit, 6-month-old kit and recently produced kit should be evaluated. For an initial (SLV), accelerated aging may be used if kits at the end of their shelf life are not available - if this is done, then lot-to-lot stability should still be performed across three recent lots. Kits should be aged using increased temperature storage as described in **ASTM F1980-16** (11) or **CLSI EP25-A** (12). Real time data is needed for validations such as AOAC Official Method applications, and prior to the first AOAC PTM renewal.

If conducted separately from the matrix/intermediate precision studies, test materials used in the evaluation should be made in any one matrix claimed for the method, using all claimed gluten sources, or using stable control materials, as long as these go through the entire testing process from extraction to interpretation. Test materials should consist of a blank, as well as a test material spiked at three times the LOQ of the method (as long as that is equal to or below 20 mg/kg, otherwise test at 20 mg/kg). Five test portions should be analyzed for each test material in each of the three kit lots.

Results should be analyzed to determine mean results, repeatability standard deviation, and recovery for each lot. These estimates must all meet acceptance criteria for all lots tested. If product stability and consistency are included in a nested design for the matrix study, data should be analyzed according to the ANOVA procedure outlined in **Annex D**.

4.8 Method Instructions and Required Method Information

Following the validation studies, the method developer should finalize the method instructions, taking into account any information learned from the validation. If detailed method preparation techniques are perceived to be proprietary information, requests may be made to the reviewers (ERP or other volunteer experts) to keep this information confidential.

Within the method instructions, the method developer must provide:

(a) A statement of the expected context(s) of use, expected matrices and expected analytical goals of the method.

(b) Specific qualifications or training required to perform the method.

(c) An applicability statement describing the method's target analyte, measurand, matrices within scope, and important limitations.

(d) If the method is intended to conform to an existing SMPR document, the SMPR citation must be provided.

(e) Step-by-step instructions for test portion preparation and performance of the method are required. Pictorial examples are encouraged.

(f) The reporting unit for all methods should be in mg/kg of gluten, although other reporting units may also be included (e.g., mg/kg of gliadin) with conversion factors.

(g) In addition to the information described in this document, method submissions must provide any additional details mandated by relevant SMPRs.

In the validation study report, method developers must provide:

(a) Information on which gluten fractions from each claimed gluten source (e.g., gliadins from wheat, hordeins from barley) the antibody/antibodies detect. Information on specific proteins or epitopes may also be provided if available.

(b) Information on calibrants:

(1) Identification of the calibrant for the method

(2) How the calibrant was prepared

(3) How the concentration value of the calibrant was assigned

(4) Whether the calibrant made from raw or processed material

(5) Whether the calibrant was extracted or purified, and the method

(6) Whether the calibrant is provided in extraction or dilution buffer

(7) How the concentration of the calibrant is expressed

(8) Whether the calibrant is commercially available.

(c) Complete information on the gluten sources (genus and species), matrices, and procedures used to prepare validation test materials.

5. Independent Laboratory Study

5.1 Scope

The independent laboratory validation study should verify the analytical results obtained in the method developer study in a controlled laboratory setting. The independent laboratory should verify the repeatability, intermediate precision, LOD/LOQ, and recovery performance parameters of the method.

5.2 Matrix Study

Incurred test materials are required for evaluation of repeatability, intermediate precision, LOD/LOQ, and recovery. *See* **Annex B** for description of best practices for incurred test material preparation.

At minimum, the independent laboratory must analyze at least one matrix for every five matrices evaluated in the Method Developer Study (Table 4), following the rotation of claimed gluten sources shown in Tables 1 or 2, depending on the method claims. The independent laboratory must analyze at least one environmental surface/Clean-In Place (CIP) solution for every five claimed. If both environmental surfaces and CIP solutions are claimed as matrices, and only one is to be included in the independent laboratory study, the environmental surface should be the chosen matrix. The selection of which matrices/surfaces/solutions are analyzed should be reflective of the range of difficulty associated with the claimed matrices.

The study design, data analysis, and reporting for the independent laboratory study should follow the same requirements described in the *Matrix Study* section of the *Method Developer Study* (**4.4**).

6. Collaborative (Interlaboratory) Study

6.1 Scope

The intent of a collaborative study is to establish relevant method attribute estimates that can be expected when a method is used in practice, with a particular focus on precision (repeatability and reproducibility) and recovery. Estimation of LOD and LOQ is also within the study scope.

Method developers may provide training on the test method to collaborator sites.

6.2 Number of Laboratories

Based on AOAC Appendix D (13) guidelines, studies must have a minimum of eight laboratories submitting valid data (to avoid unduly large confidence bands about the estimated parameters).

To minimize potential bias, no more than 25% of the laboratories with data included in the final dataset may come from the same organization. For this purpose, the term "organization" includes companies (test kit manufacturers, method developers, food processors, etc.), regulatory bodies, government agencies, or any other body (Appendix M; 14)

6.3 Matrix Study

The collaborator sites will perform the matrix studies for each claimed gluten source in at least one of the incurred matrices for each matrix category claimed in the method developer study, following the rotation of claimed gluten sources shown in Tables 1 or 2, depending on the method claims. The selection of which matrices/surfaces/solutions are analyzed should be reflective of the range of difficulty associated with the claimed matrices.

If the method developer study consisted of only individual matrices, rather than matrix categories, then the collaborator study will test at least one incurred matrix for every five matrices tested in the method developer study, as shown in Table 4. The collaborator sites must also analyze at least one environmental surface/CIP solution for every five claimed. If both environmental surfaces and CIP solutions are claimed as matrices, and only one is to be included in the collaborative study, the environmental surface should be the chosen matrix.

Table 4. Number of matrices to be tested by each independent or collaborator site, as related to the number of claimed matrices.

Claimed matrices	1-5	6-10	11-15	16-20
Matrices tested by independent or collaborator labs	1	2	3	4

The selection of the specific matrices used in the collaborative studies should be reflective of the range of difficulty and matrix category associated with the claimed matrices.

6.4 Test Materials

Appendix D requires a minimum of five materials be used in the collaborative study (13). Each claimed matrix should be tested with at least one gluten source (per Tables 1 and 2) at a minimum of four concentration levels, including zero.

Two blind-coded replicate test portions should be analyzed by each laboratory for each test material (i.e., each matrix-concentration combination). For each matrix, the concentration levels must include a blank (zero) and a level at less than or equal to two times the LOQ stated in the kit insert (as long as this is less than or equal to 20 mg/kg, otherwise test at 20 mg/kg). The remaining concentrations should be distributed throughout the quantification range. (14)

Incurred test materials are required for estimation of precision, LOD/LOQ, and recovery. *See* **Annex B** for description of best practices for incurred matrix preparation.

6.5 Data Analysis

All individual data values must be reported.

Data analysis will be conducted according to the procedures described in Appendix D (13). Specifically, the following must be performed and reported:

(1) Outliers should be evaluated as described in Appendix D (13).

(2) Recovery must be reported, with calculations using the known quantity of target present in incurred test materials based on gravimetric calculations and accounting for any mass balance changes occurring during food processing (e.g., moisture loss during baking).

(3) Precision estimates reported must include both repeatability (S_r and RSD_r) and reproducibility (S_R and RSD_R).

(4) LOD/LOQ.—LOD and LOQ will be estimated using reproducibility data (S_R and RSD_R).

Data collected from analysis of incurred test materials for all matrices will be used to model the relationship between analyte concentration and reproducibility. Data used must meet other method performance criteria (e.g., recovery).

LOD will be estimated using a hypothesis test approach, with $\alpha = \beta = 0.05$. The relationship between observed concentration and intermediate precision standard deviation must be taken into account in the estimation of LOD (also referred to as a precision profile estimation method for LOD). Full instructions for the calculations to estimate LOD are in **Annex D**.

LOQ estimation will be based on the relationship between concentration and intermediate precision standard deviation. Full instructions for the calculations to estimate LOQ are in **Annex D.**

6.6 Acceptance Criteria

All parameter point estimates must meet any applicable requirements for confidence intervals established by the AOAC Statistics Committee or other relevant reviewers.

If an applicable SMPR is available for a method, the collaborative study data must meet the corresponding criteria.

In the absence of an applicable SMPR, an expert review panel will evaluate the study data according to their expert opinions.

(a) *LOQ*.—The LOQ must be greater than or equal to the LOD.

The RSD_R at the LOQ must be less than or equal to the RSD_R in the relevant SMPR. If there is no SMPR available, RSD_R at the LOQ must be $\leq 30\%$.

If a method developer has an LOQ claimed as part of the method design (e.g., the lowest non-zero calibrant), the estimated LOQ from the Collaborative Study (which meets the SMPR requirements for maximum RSD_R) must be less than or equal to the claimed LOQ of the kit, within statistical tolerances. If the estimated LOQ from the Collaborative Study is greater than the claimed LOQ of the kit, the method developer must revise the LOQ claimed in the test kit insert and validation reports to meet the precision requirements for LOQ.

In the validation reports and test kit inserts, the method developers must indicate the actual RSD_R value estimated for the LOQ of the kit as part of the LOQ information. For example:

LOQ_{15}, for a method where the existing LOQ claimed by the kit had an estimated RSD_R of 15% in the Collaborative Study

LOQ_{30}, for a method where the LOQ was set based on the Collaborative Study outcome and a maximum RSD_R of 30%

Acceptance criteria for the maximum RSD also includes meeting requirements for confidence intervals, as established by the AOAC Statistics Committee.

6.7 Collaborator Comments

Comments on the candidate method should be encouraged from all collaborators, and any comments should be reported in the collaborative study report.

7. Matrix Extension

7.1 Matrix Extension for SLV Studies

A single laboratory matrix study must be performed as described in *4.4*, picking up where the laboratory left off in the gluten source tables (Tables 1 and 2). A matrix study must also be completed by an independent laboratory, and reported, as described under **5.2**.

7.2 Matrix Extension for Multi-Site Collaborative Studies

A single laboratory matrix study must be performed as described in **4.4**, picking up where the laboratory left off in the gluten source tables (s*ee* Tables1 and 2). A minimum of eight collaborator sites will perform the matrix studies as described under **6.3** and reported as described under **6.5**.

Documents Consulted

AAFCO Good Samples and Good Test Portions: https://www.aafco.org/resources/guides-and-manuals/good-test-portions-and-goodsamples-resources/

FDA ORA-LAB 5.4.5 (2023) Volume II —Methods, Method Verification and Validation, Document No IV-02, Version 2, Section 2—Microbiology.

ISO/IEC Guide 99:2007, International vocabulary of metrology—Basic and general concepts and associated terms (VIM)

Koerner et al. (2013) JAOAC 96 (5), 1033-1040.

USP 31:2008, U.S. Pharmacopeia General Information/ Validation of Alternative Microbiological Methods

References

(1) De Bièvre, P. (2013) *Accreditation and Quality Assurance* **18**, 71-72

(2) CODEX STAN 118-1979: *Standard for foods for special dietary use for persons intolerant to gluten*

(3) ISO 3534-2:2006, *Statistics — Vocabulary and symbols—Part 2: Applied statistics*, International Organization for Standardization, https://www.iso.org/standard/40147.html

(4) ISO 5725-1:2023, *Accuracy (trueness and precision) of measurement methods and results—Part I: General principles and definitions*, International Organization for Standardization, https://www.iso.org/standard/69418.html

(5) ISO Standard 17511 (2020) Reference Materials Selected Terms and Definitions, International Organization for Standardization, https://www.iso.org/standard/69984.html

(6) CLSI (2018) EP07-A2, *Interference Testing in Clinical Chemistry*, in *CLSI Guideline EP07-A2*, Clinical and Laboratory Standards Institute, Wayne, PA, USA

(7) ISO 16577:2022, *Molecular Biomarker Analysis*-Vocabulary for Molecular Biomarker Analytical Methods in Agriculture and Food Production, International Organization for Standardization, https://www.iso.org/standard/81024.html

(8) *Official Methods of Analysis of AOAC INTERNATIONAL* (2023) Appendix F: Guidelines for *Standard Method Performance Requirements*, G.W. Latimer, Jr. (Ed.), Oxford University Press, New York, NY, USA. https://doi.org/10.1093/9780197610145.005.006

(9) Abbott, M., Hayward, S., Ross, W., Godefroy, S.B., Ulberth, F., van Hengel, A.J., Roberts, J., Akiyama, H., Pöpping, B., Yeung, J.M., Wehling, P., Taylor, S.L., Poms, R.E., & Delahaut, P. (2010) *J. AOAC Int.* **93**, 442–450

(10) AOAC SMPR 2017.021 (2017) Standard Method Performance Requirements (SMPRs) for Quantitation of Wheat, Rye, and Barley Gluten in Oats G.W. Latimer, Jr. (Ed.), Oxford University Press, New York, NY, USA. https://doi.org/10.1093/9780197610145.003.1134

(11) ASTM F1980-16 (2016) Standard Guide for Accelerated Aging of Sterile Barrier Systems for Medical Devices. https://www.astm.org/f1980-16.html

(12) CLSI (2009) EP25-A, *Evaluation of Stability of In Vitro Diagnostic Reagents; Approved Guideline*, Clinical and Laboratory Standards Institute, Wayne, PA, USA

(13) *Official Methods of Analysis of AOAC INTERNATIONAL* (2023) Appendix D*:* Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis, G.W. Latimer, Jr. (Ed.), Oxford University Press, New York, NY, USA. https://doi.org/10.1093/9780197610145.005.004

(14) *Official Methods of Analysis of AOAC INTERNATIONAL* (2023) Appendix M: Guidance on Food Allergen Immunoassay Validation, G.W. Latimer, Jr. (Ed.), Oxford University Press, New York, NY, USA. https://doi.org/10.1093/9780197610145.001.0001

ANNEX A

Selectivity Study

Buy from reputable sources and ensure that you are getting the actual material, and that it's gluten free. This can be done by testing using an appropriate validated method.

If any of the matrices purchased as a flour are demonstrating unexpectedly high results, it may be necessary to purchase that material in a whole grain/seed/bean form and grind it in your own lab, in order to get a clean material that will give you a true estimation of cross-reactivity.

If you have information on the specific varietal tested, include that information in the validation report, as well as including the part(s) of the material that is tested (skin, flesh, stone, pit, etc.). For a multi-component matrix like pork sausage, provide all ingredients.

Table A1. Gluten sources (materials should be tested at three times the LOQ, as long as that is equal to or below 20 mg/kg, in rice flour)
Wheat flour[a] (*Triticum aestivum*)
Wheat flour[b] (*Triticum compactum*)
Durum wheat flour[b] (*Triticum durum*)
Einkorn wheat flour[b] (*Triticum monococcum*)
Emmer wheat flour[b] (*Triticum dicoccon*)
Khorasan wheat flour[b] (*Triticum turanicum*)
Spelt wheat flour[b] (*Triticum spelta*)
Triticale flour[b] (*Triticosecale*)
Oat flour[a,c,d] (*Avena sativa*)
Rye flour[a] (*Secale cereale*)
Barley flour[a] (*Hordeum vulgare*)

[a] These may be omitted if they are being used as a gluten source in the validation matrix studies.
[b] For all minor wheat species and Triticale, 20 ppm samples can be prepared using the protein-to-gluten conversion factor in Annex B, or the method developer may use the cited wet chemistry method to determine their own conversion factor. All methodology and findings must be included in the study report.
[c] Oats are not regulated as a gluten source in all countries. *See* the definition of "Gluten", and the related footnote, in the main guidance document. If oats are not considered a gluten source for the test method undergoing validation, then oats should be moved to Table 2 and treated as a commodity for the cross-reactivity and interference studies.
[d] Oats that are not comingled with wheat, rye or barley may be difficult to source. Whole oat groats may need to be ground to generate a pure oat flour sample.

14

15

Table A2. Commodities for cross-reactivity and interference studies (materials should be tested as normally purchased/used. Any processing should be described (roasting, irradiation, etc.)[a]
Almond flour (*Prunus dulcis*)
Amaranth flour (*Amaranthus* spp.)
Arrowroot (*Maranta arundinacea*)
Black bean flour (*Phaseolus vulgaris*)
Beef meat (*Bos taurus*)
Brown rice flour (*Oryza sativa*)
Buckwheat flour (*Fagopyrum esculentum*)
Carob (*Ceratonia siliqua*)
Chestnut flour (*Castanea sativa*)
Chicken meat (*Gallus gallus domesticus*)
Cocoa
Coconut flour (*Cocos nucifera*)
Ground coffee (*Coffea arabica* or *Coffea canephora*)
Corn meal (*Zea mays*)
Dried fruits or raisins (*Vitis vinifera*)
Egg powder, chicken (*Gallus gallus domesticus*)
Faba bean flour (*Vicia faba*)
Flax seed flour/ meal (*Linum usitatissimum*)
Garbanzo bean/chickpea flour (*Cicer arietinum*)
Green pea flour (*Pisum sativum*)
Guar gum, dilute 1:10 in rice flour (*Cyamopsis tetragonoloba*)
Hazelnut flour (*Corylus avellana*)
Lentil flour (*Lens culinaris*)

Lima bean flour (*Phaseolus lunatus*)
Lupin Flour (*Lupinus spp*.)
Milk powder, cow (*Bos taurus*)
Millet flour (*Panicum miliaceum*)
Oat flour[b,c] (*Avena sativa*), if not a claimed gluten source
Pea protein (*Pisum sativum*)
Peanuts (*Arachis hypogaea*)
Pork sausage (*Sus domesticus*)
Potato flour/starch (*Solanum tuberosum*)
Quinoa flour (*Chenopodium quinoa*)
Salmon (*Oncorhynchus* spp.)
Sesame flour (*Sesamum indicum*)
Sorghum flour (*Sorghum bicolor*)
Soya flour (*Glycine max*)
Spices (at least one from Table A3)
Sweet rice flour (*Oryza sativa glutinosa*)
Tapioca flour/starch (*Manihot esculenta*)
Tea, ground (*Camellia sinensis*)
teff flour (*Eragrostis tef)*
Walnuts (*Juglans* spp.)
White bean flour (*Phaseolus vulgaris var. humilis*)
White rice flour (*Oryza sativa*)
Yellow pea flour (*Lathyrus aphaca*)
Xanthan gum, dilute 1:10 in rice flour (from *Xanthomonas campestris*)

[a] Adapted from Koerner et al. (2013) *J. AOAC Int.* **96**, 1033–1040.

[b] Oats are not regulated as a gluten source in all countries. *See* the definition of "Gluten", and the related footnote, in the main guidance document. If oats are not considered a gluten source for the test method undergoing validation, then oats should be moved to Table 2 and treated as a commodity for the cross-reactivity and interference studies.
[c] Oats that are not comingled with wheat, rye or barley may be difficult to source. Whole oat groats may need to be ground to generate a pure oat flour sample

16
17

Table A3. Possible additional commodities (materials should be tested as normally purchased/used - any processing should be described (roasting, irradiation, etc.)	18
Carrageenan (dilute 1:10 in rice flour)	
Cauliflower *(Brassica oleracea var. botrytis*)	
Chia (*Salvia hispanica*)	
Cinnamon (*Cinnamomum verum*)	
Clove (*Syzygium aromaticum*)	
Coriander seed (*Coriandrum sativum*)	
Cumin (*Cuminum cyminum*)	
Ginger powder (*Zingiber officinale*)	
Hemp (*Cannabis sativa*)	
Kidney bean flour (*Phaseolus vulgaris*)	
Marjoram (*Origanum majorana*)	
Paprika (*Capsicum annuum*)	
Parsley flakes (*Petroselinum crispum*)	
Poppy Seeds (*Papaver spp*.)	
Protein sources (e.g., Duckweed (*Lemna minor*), insect, algal, fungal	
Rye grass (*Lolium perenne*)[a]	
Romano bean flour (*Phaseolus coccineus*)	
Sage (*Salvia officinalis*)	
Sunflower kernels (*Helianthus annuus*)	
Thyme (*Thymus vulgaris*)	
Turmeric (*Curcuma longa*)	
Urad Dal flour (*Vigna mungo*)	

[a] Subject to further research, this may be of interest as a gluten-like source.

ANNEX B

Preparation of Materials for Gluten Method Validation

Until such time as reference materials are available, the gluten source for all prepared samples should be commercial, unbleached whole wheat, whole rye, whole barley, or whole oat flour.

The chosen flour should be analyzed for Dumas or Kjeldahl nitrogen. Convert to percent crude protein by multiplying the nitrogen value by 5.83. Then convert to percent gluten by multiplying the crude protein value by the following factors, depending on the grain:

Wheat	0.74
Rye	0.52
Barley	0.78
Oats	0.15

These conversion factors are suggestions and may vary across different grain samples. The factors come from two publications (1, 2); the conversion factors for wheat rye and barley are based on the wet chemistry method described in Wehling and Scherf (2). Method developers may also use the wet chemical method in Wehling and Scherf (2) to arrive at the gluten content for their own wheat, rye and barley flours.

Finally, convert the percent gluten to mg/kg (ppm) gluten by multiplying the result by 10,000.

As an example, a barley flour is tested and found to have a Dumas nitrogen level of 1.5%. This is multiplied by 5.83 to attain a crude protein level of 8.75%. Using the conversion factor for barley, the 8.75% crude protein is multiplied by 0.78 to obtain the gluten percent of 6.825%. This percent value is then multiplied by 10,000 to estimate the mg/kg (ppm) value at 68,250. This is equivalent to 68.3 mg of gluten per gram of flour.

B1. Making Spiked Materials

Bulk spiked materials may be prepared for the selectivity, stability and lot-to-lot studies, and bulk spikes of raw materials are often made prior to the processing steps when making incurred samples.

These methods can be used for any material that has a small particle size or uniform consistency, including flours, baking mixes, spices, meats, sauces, dressings, ice cream (melted), etc. They can also be used in other matrices that can be dried and ground to a flour-like consistency, such as nuts, seeds, and breadcrumbs.

Thorough blending is key to a successful trial. For dry materials like flours, or for liquid consistencies, blending can be done in a blender or tumbler-style mixer, or even by manual tumbling of material in a zippered plastic bag. Add the spike material uniformly within the

matrix, rather than adding it all in one location prior to blending it in. Making spikes in very fine matrices with small particle size can be difficult, and re-milling of the matrix and spike may be necessary to achieve particle size homogeneity. While gluten is not water-soluble, it can be uniformly dispersed in sauces, dressings, and other liquids by either spiking directly with flour, or making a suspension of gluten in the matrix, mixing it thoroughly to achieve uniformity, and using this to make the spikes. Make sure to mix the material again before any samples are taken from it. For paste-like items and meats, spread the matrix out on aluminum foil, parchment, or other non-stick surface, sprinkle the spike material uniformly across the top, and then recombine the matrix and mix by kneading. Extremely high-speed or high-heat mixing can alter the gluten results, so mechanical blending should be done in short pulses, and only for the duration needed to achieve sufficient uniformity.

Liquid suspensions made in the kit extraction buffer can be used to spike individual test portions for the interference portion of the selectivity study prior to extraction. Liquid spiking of test portions may not be used for the matrix or other studies. If this method is used, state in the validation report that the selectivity study only tests for analytical interference, not interference with the extraction.

Options for adding gluten to the matrix, either as a spike or prior to processing of an incurred matrix, include (*see* Figure B1):

1. Creation of a mid or high-level stock followed by serial dilution. The gluten concentration in the stock should be chosen to allow the largest volume of stock material to be used in the preparation of each spike level.
2. Creation of mid or high-level stock used to then make each individual bulk preparation.
3. Creating bulk spike level samples directly from the source material (flour).
4. A combination of the above, in which spikes are made directly from the flour source for higher levels, then diluted to achieve lower levels).

The method for creation of each sample must be described in the report.

Any suitable validated quantitative method can be used to assess sample homogeneity. Assessing homogeneity of the high or mid-level stock can be a good initial step before preparing lower-level spikes. Homogeneity should be assessed for every bulk test material, or at least as many as needed to confirm that the mixing procedure is adequate to minimize distributional variance.

Homogeneity should be assessed by testing 10 test portions, taken from throughout the material, individually extracted, and run according to the method instructions of any validated quantitative assay (e.g., use 2 wells if the method calls for it).

The preferred CV from the homogeneity data will depend on the method performance requirements, with the homogeneity SD below the required repeatability SD. Higher CVs may be expected at lower analyte concentrations.

Use the stocks for testing on the same day if possible. Samples made in dry matrices, like flours, can be stored at room temperature for several days, remixing each stock thoroughly before use. Samples made in perishable matrices (dairy products, meats) should be refrigerated for no more than 2 days, remixing each stock thoroughly before use. Samples may also be stored frozen in working aliquot-sized portions for an extended period.

B2. Making Incurred Materials

The section above, *Making Spiked Materials*, describes the initial steps in making an incurred material. The spiking must occur prior to the major processing step in order for the end product to be considered an incurred matrix. Further considerations for common types of processing are provided below

B2.2.1 Baked, Fried, or Dehydrated Materials

Baking, frying, and dehydrating are processing methods that can be reasonably replicated at a small scale, in a laboratory. The same process applies for each.

When possible, weigh the incurred material before and after processing. Any change in the analyte concentration above or below the expected value should be accounted for by the change in mass.

When exact ppm values are needed, for example for a quantitative method, the moisture/weight change from processing must be accounted for in determining the amount of spike material to be added. If the entirety of the material cannot be weighed before and after processing, additional analyses can be performed to determine the potential analyte gain or loss, such as moisture content, protein, or zinc/other metals.

If moisture/weight change results in a slightly higher ppm value than intended, higher-level incurred samples can be mixed with blank, processed sample to achieve various concentrations. The lowest concentration achieved in this way should not be less than 10% of the concentration of the high-level incurred material. Larger discrepancies require a second incurred matrix to be made at a lower level.

B2.2.2 Pressure Treated/High Heat/Extruded

These are processes that cannot normally be replicated outside of a manufacturing facility. If a manufacturer is particularly interested in the development of the assay, the kit developer may be able to partner with them to make gluten spikes on a pilot scale, using a similar method as described above for baked, fried and dehydrated products.

In the absence of access to a manufacturing plant, some highly processed matrices can be incurred through “fortification”. An example would be a whole wheat puffed/extruded breakfast cereal. A pilot plant could create a mid-level spike (100 ppm, for example), which could be diluted down in a similarly processed blank matrix to create lower concentrations.

Any validated method can be used to verify the absence of gluten in the non-gluten-containing matrix.

B3. Making Environmental Surface Samples

Determine the surface area that’s expected to be swabbed. Typical area is 25 cm^2 – 100 cm^2 (approx. 4 $in.^2$ – 16 $in.^2$)

Make suspensions from the flour in the kit extraction solution, or 60% ethanol solution.

Create solutions at gluten concentrations (µg/mL) around the expected sensitivity level of the method, as described in the validation requirements.

Pipette gluten suspension per outlined area, distributing the liquid as evenly as possible. Shake the suspension thoroughly before pipetting it into each square. Note the volume of solution added to each area, to allow the total µg of gluten per swab area to be calculated.

If the method is for swabbing of wet areas, the surfaces are ready for testing. If the method is meant to test dried-on material, allow the gluten suspension to dry completely (overnight if necessary).

Cleaning solution studies for an environmental surface claim are voluntary. Cleaning solution studies are to be performed as described in the following section.

B4. Making Rinse Water/CIP/Cleaning Solution Study Samples

Make a high-level suspension of gluten in kit extraction buffer or 60% ethanol, then dilute into water or water/cleaning solution to the desired gluten concentrations, around the expected sensitivity level of the method, as described in the validation requirements. The high-level gluten solution in kit extraction buffer or ethanol should not make up more than 1% of the final CIP (cleaning solution)/rinse water preparation, to ensure that the sample is representative of a typical CIP (cleaning solution)/rinse water sample. If the method is designed for rinse water testing, and the cleaning solution will not be tested at its recommended working concentration, the dilution of the cleaning solution must be reported.

Cleaning solution (CIP) validations must be performed separately for each cleaning agent. Method developers may choose to perform a validation in examples from each of the four main types of cleaning solutions: degreasers, detergents, abrasives, and acids. But the validation will only be reported for the specific cleaning agent that is used.

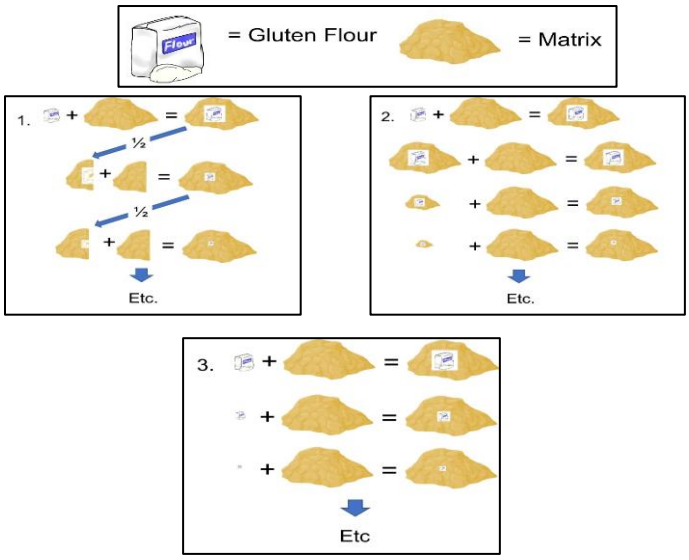


Figure B1. Options for generating bulk spike materials at various levels.

References (Annex B)

(1) Schalk K., Lexhaller B., Koehler P., & Scherf K.A. (2017) *PLoS ONE* **12**, e0172819. doi:10.1371/journal.pone.0172819

(2) Wehling, P. & Scherf, K.A. (2020) *J. AOAC Int.* **103**, 210-215

Annex C

AOAC Triangle was reviewed and considered not to be relevant for gluten analysis.

For each material tested, specifics of the formulation and processing, including percent fat, protein and pH must be provided.

Refer to Codex food standard definitions http://www.fao.org/gsfaonline/foods/index.

FDA list https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=FoodSubstances.

Table C1.

	Not fermented, hydrolyzed or fractionated							
Food category	Raw/minimally processed	Processed/baked/ cured/smoked/ marinated	Fried	Pressure/heat-UHT/ pasteurization/ extrusion	Dehydrated/dried/ dry cured	Fermented	Hydrolyzed	Fractionated
	Raw/minimally processed					Fermented	Hydrolyzed	Fractionated
Binders, stabilizers, emulsifiers	Carrageenan					Xanthan gum, guar gum		
		Processed						
Candy		Caramel, pralines, marzipan, nougats, Pastilles, Lozenges, jelly beans, toffees, licorice, Chewing gum, Mints, Icing or Frosting (non-chocolate), sauces used for toppings, non-chocolate (butterscotch, marshmallow)						
	Raw/minimally processed	Processed, baked	Fried	Pressure/extrusion	Dehydrated	Fermented	Hydrolyzed	Fractionated
Cereal grains	Whole or milled Sorghum, soybeans, corn, millet, teff, rice, fonio, oats; baking mixes	Bread, cakes, cookies, tortillas, fresh pasta, bakery products, confectionaries, crackers, bagels, muffins, grain-based protein bars	Breaders/Batters for fish sticks and chicken nuggets, tortilla chips, donuts	Breakfast cereals, puffs/pellets	Breadcrumbs, dried pasta	Sourdough, malt, malt extract, sprouted flours, soy, oat or rice based yogurts, natto, tempeh, soy sauce, miso	Soy, oat, rice, or teff beverages (if hydrolyzed as part of processing); modified food starch.	Unmodified wheat starch, soy protein isolate, tofu, maltodextrin, soy lecithin

	Raw/minimally processed			
Chemicals and preservatives	Need to be validated per matrix			
	Raw/minimally processed	Processed		
Chocolate and cocoa	Cocoa nibs, cocoa mass,	Cocoa powder, cocoa sugar mixes, baking chocolate, chocolate chips, chocolate bars, chocolate syrup, cocoa butter, chocolate-hazelnut spread, chocolate frosting, enrobing chocolate, and similar carob-based products		
Cleaning solutions and rinses	Rinsates, CIP rinse - validate per cleaning solution			
	Raw/minimally processed		Dried	
Colors, flavors, and fragrances (other than those prepared as extracts)	Caramel color, caramel flavor, liquid smoke, beverage flavorings for water, coffee		Spirulina, beetroot powder, pitaya powder, acai berry powder, matcha green tea powder, turmeric, saffron, annatto extract	
				Fermented
Cultured materials				Validate per culture strain
	Raw/minimally processed	Processed	Pasteurized/ heated/UHT	Fermented
Dairy products (high fat)	Raw butter, raw cream	Whipped cream, cream cheese, custards, puddings	Ice cream, canned creams	Raw or pasteurized greek yogurt, skyr, semi-hard cheeses (e.g., comte, beaufort), blue cheese (roquefort), soft cheese (e.g., brie, munster)

Dairy products (not high-fat)	Raw milk (skim, part skim or whole), raw dairy-based drinks		Pasteurized dairy products, UHT milks, canned milks, pasteurized dairy-based drinks (chocolate milk, strawberry milk), ice milk	Condensed milk, evaporated milk, sweetened condensed milk, blends of condensed milk and vegetable fat (liquid or powdered beverage whiteners), khoa, milk powders, powder for milk-based desserts	Raw or pasteurized fermented milk, yogurt, buttermilk (dried, liquid), clotted cream, kefir, sour cream, cheese spray (dried/liquid), flavored yogurt drinks, lactic acid bacteria drinks (lassi), junket	Raw or pasteurized acidified milk, yogurt, buttermilk (dried, liquid), cheese spray (dried/liquid), chhena	Whey concentrate or isolate, casein, milk protein concentrate or isolate, whey-based drinks, whey cheese, ghee, anhydrous milkfat, anhydrous butter oil, butter acids
		Processed		Dried	Fermented		
Dressings, condiments and marinades		Mayonnaise, mustard, pesto, ketchup, sauces, salad dressings, marinades, onion dip, chili sauce, sweet and sour sauce, barbecue sauce, cheese sauce, white sauce, gravies, oyster sauce, fish sauce		Sauce and gravy mixes	Vinegars, soy sauce		
	Raw/ minimally processed	Processed	Pasteurized	Dried			Fractionated
Eggs	Shell eggs, whole eggs, egg yolks, egg whites	Quiche	Pasteurized whole egg pulp, Pasteurized egg whites	Dried whole egg, dried egg whites			Egg protein
Environmental surface swabs	Per surface						
						Hydrolyzed	
Enzymes						Validated per enzyme	
							Fractionated
Extracts							Acai berry extract, aloe extract, vanilla extract
	Raw/minimally processed	Processed	Heat treated			Hydrolyzed	Fractionated

Fats, oils and fat emulsions	Virgin or cold-pressed olive oil, cottonseed, oil, peanut oil, vanaspati, almond oil, apricot kernel oil, coconut oil	Margarine, butterine, minarine	Lard, rendered animal fats, fish oils, tallow			Mono- and diglycerides	Wheat germ oil, basil oil, bergamot oil, carrot oil

	Raw/minimally processed	Processed	Heat treated	Dried
Fish - shellfish - seafood	Fish, oysters, clams, scallops, mussels, shelled, shrimp	Frozen fish sticks, baked fish, Shelled and shucked products of cooked crustaceans, fish and seafood, roll herring, anchovies, smoked fish, imitation crab, imitation lobster	Canned fish, canned crab	Dried (salted) fish, bonito (dried)

	Raw/minimally processed	Processed	Pasteurized/heat treated	Dried	Fermented	Fractionated
Fruits and vegetables (*see* colors category for fruits/vegetables used as coloring - *see* nutraceuticals for concentrates used as supplements)	Any raw or frozen fruits, Fruit mixes, potatoes, yams, sweet potatoes, cassava, dahlia, carrots, cruciferous vegetables, sprouts (eg alfalfa, soy, fenugreek, mung), fresh herbs (eg basil, cilantro, parsley); cassava flour, konjac flour, sago flour, bagged pre-cut leafy vegetables, salad mixes, shredded or riced vegetables	Juices, smoothies, blanched vegetable, fruit 'leather'	Pasteurized juices, canned juices, jams, jellies, marmalades, canned fruits and vegetables, apple butter, lemon curd, mango chutney, raisin chutney, candied fruit (e.g., maraschino cherries, candied citrus peel), fruit preparations (as for yogurt)	Dried fruits (not coated in flour), air dried or freeze dried vegetable snacks (e.g., peas, chickpeas, sweet potato chips, beet root chips)	Fermented cabbage, pickles, pickled plums, mango pickles, pickled gooseberry	Syrups, inulin

			Pasteurized/heat/ UHT	Dried	Fermented	Hydrolyzed	Fractionated
Infant formula			FSMPs and liquid IMF (milk and plant based) high fat or high protein (for FSMPs)	Dehydrated milk, dehydrated yogurt	Spray-dried pre-blend, culture powders	Soy-based and plant-based formulas, milk-based formula and FSMP, amino-based IMFs	Probiotic whey-based formula, probiotic soy-based formula

	Raw/minimally processed	Processed	Pasteurized/ heat treated	Dried	Fermented

Juices, soft drinks, beverages, edible ices	Sorbet	Sport drinks, energy drinks, carbonated drinks		Fruit juices, tomato juice	Juice concentrate, tea (dried), ground coffee	Beer, kombucha, wine, cider, perry, mead, spirits, sake, malt based coolers, seltzer	Prepared coffee, prepared tea
	Raw/minimally processed				Dried		Fractionated
Legumes	Peanut, chickpea (garbanzo), pea, beechnut, pigeon pea, soybean, lima, adzuki, black, mung, cowpea, lentil, beans (fava, kidney, pinto, white), carob, lupin, mesquite				Pulses (dried beans, chickpeas, lentils, peas, soybeans); legume flours		Pea protein

	Raw/minimally processed	Processed/cured/ smoked/marinated	Fried	Heat treated	Dried	Fermented	Fractionated
Meat, game, and poultry	Carcasses, meat cuts, carpaccio's, minced meat, meat preparations ready to cook, frozen burger patties, marinated beef shish-kebab, seasoned portions	Roast beef, brisket, sausages, cooked ham, pate, cooked meats, cured meats (filet de sax), cooked filets, smoked poultry, corned beef, sausages, lunch meat, hot dogs, bologna	Par-fried (frozen) meat patties, Fried bologna	Canned poultry, Canned poultry pate	Jerky, cobourg ham, dry cured ham	Salami	Fish protein isolate, collagen, gelatin
	Raw/minimally processed				Dried		Fractionated
Non-seed crop materials	Barley grass juice, wheat grass juice,				Alfalfa powder, barley grass powder, wheat grass powder, oat hulls, rice hulls, dried algaes		Arrowroot starch, tapioca starch
						Fermented	
Nutraceuticals – dietary supplements Pet food and animal feed						Vitamins, amino acids	
						Fermented	

Probiotics					Validated per bacterial strain		
	Raw/minimally processed						Fractionated
Pseudocereals	Amaranth, goosefoot, kaniwa, quinoa, hanza, chia, flax (linseed), breadnut, sesame, buckwheat						Flax protein, chia powder, hemp protein
				Dried			Fractionated
Seeds				Hemp, mustard, poppy, sunflower, pumpkin, pomegranate			Psyllium fiber, hemp protein
			Heat treated/UHT	Dehydrated			
Soups, broths, stocks			Bouillon, broth, consommé, water and cream-based soups, chowder, bisque, canned or bottled, aseptically packaged soups, broths, stocks (Tetrapaks)	Bouillon powder, cubed stocks, powdered soups, condensed soups			
	Raw/minimally processed			Dried			Fractionated
Spices	Salt, salt substitutes, onion salt, garlic salt			All dried spices (allspice, caraway, cardamom, celery seed, cinnamon, cumin, fenugreek			Allspice oleoresin, cardamom oleoresin, celery seed oleoresin
	Raw/minimally processed	Processed	Heat treated	Dehydrated		Hydrolyzed	Fractionated
Sugars/ sweeteners	Honey	Cane juice, xylitol, aspartame, saccharin, sucralose	Candy syrups	Coconut sugar, palm sugar, molasses, stevia		Invert sugar	Rice syrup
	Raw/minimally processed	Processed	Heat treated/UHT			Hydrolyzed	
Tree nuts (combine with seeds?)	Pecans, walnuts, almonds, hazelnuts, cashews	Almond butter, nougat, pecan shell flour, almond flour	Coconut milk, coconut cream, nut-based milks, whipped toppings, creamers, sour cream substitutes			Nut-based milks, whipped toppings, creamers, sour cream substitutes that have been enzymatically treated	
					Fermented		Fractionated

Yeast		Yeast (bakers, nutritional, torula), yeast extracts, koji (Validated per yeast strain)	Baker's yeast glycan, baker's yeast protein

ANNEX D

Statistical Methods for Quantitative Gluten Assays: Data Analysis Guidance and Example Datasets

1 Intermediate Precision and Repeatability Estimation From Nested Designs: Analysis of Nested SLV Designs In R

(Courtesy of Paul Wehling, ChemStats Consulting LLC)

1.1

As is described in the guidance, intermediate precision and repeatability can both be estimated from one of several nested designs.

1.2 Basic Principles of the Nested Designs

1.2.1 Defining the Variance Components

When validating a method with a nested experiment, it is strongly recommended that researchers define terms used to describe the experimental factors. Because all methods are different, and researchers tend to use different words to convey the same meaning, it is important to define terms in order to avoid confusion. For example, in the largest design in the Guidance, Design 2b, there are potentially 4 levels of experimental factors that can be differentiated and estimated: Lot, Analyst/Day, TP, and ELISA. Now in all designs, 1a, 1b, 2a, 2b, there is an explicit understanding that Analyst and Day are confounded and will be included in the model as a single factor. In addition, each of these levels may have many more sources of variation than just those given by the 4 terms used. It is recommended to explicitly write out the sources of variation and how they contribute to the 4 variance components that will estimated experimentally. Nested experiments are unique in this aspect. Generally, with a factorial experiment, you can control the conditions so that only the interested factors are varied.

(a) *Terminology.*—"Source of variance" refers to a specific source of variation in the method for example, weighing variation. This refers to all of the small sources of variation that add together to make the overall measurement uncertainty.

"Variance component" is a statistical term for a collection of one or more sources of variation that will be estimated by the validation experiment. In this case, we will have 4 variance components. The purpose of this exercise is to take all of the known sources of variation and assign them to one of the 4 variance components. The distribution of sources of variation depends on the experimental conditions and how the analyses were performed.

(b) *Example of variance component description for a nested experiment of a typical ELISA method.—Note*: the following are for a hypothetical ELISA method – ALL METHODS ARE UNIQUE and will be different – this should be performed for each method and each validation.

(1) Lot includes.—Manufacturing variance of the lot, potentially different response of antibodies. Certain reagents are unique to each lot, so there will be reagent variance.

(2) Analyst/day includes.—Different operators, different times, different days, different teams, different environmental conditions in the lab, DIFFERENT CALIBRATIONS on different plates, different temperature.

(3) TP includes.—Test portion variation due to sampling, heterogeneity of the analytical sample (compositional and distributional), weighing variation, volume addition variation, extraction variation: time, temperature, water bath fluctuation. This variance component will include

everything that can happen within a set from weighing of the test portion until you are ready to take the aliquot of the extract onto the ELISA plate.

(4) ELISA includes.—Aliquot variation, heterogeneity in the extract, reagent pipetting variance, differences in coating of the wells, well-to-well sensitivity variation, rinsing issues, pipetting volumes, different optical density of each well, reader issues, timing of color development, how fast you pipet from start to finish, different development times across the plate.

In order to make the software work, you need to give a name to each of the 4 variance components, with the understanding that there will be several sources of variation within each variance component category. I suppose you could call them Level 1, Level 2, Level 3, Level 4, but the usual way to do this is to take what you think is the most important source and use it as the "name" of the variance component – keeping in mind that the name is only a label and if you call the 3rd level "Test Portion" that doesn't mean that all those other sources are gone – this is just the Label we are using for convenience. (In these experiments, "Test Portion" will usually always include extraction sources as well.)

(c) It's critical to do the categorization of variance sources into variance components for two reasons: First, it is important to define terms, but more importantly, it will come in handy to determine if the factor is nested.

Note: If you only do the 3-factor experiments such as Design 1a or 1b, the variance components above labeled as "TP" and "ELISA" will be combined into 1 component. So, you may call that combined component "TP", but it will contain all of the ELISA variance sources in addition to the other sources. (Maybe "TP" is not a good name for that in the 3-factor design.) Researchers are free to use any label for the name of the variance component, but this should always be understood that there are more sources of variation within a variance component than the one that is used as the label.

1.3 What is a "Nested" Experiment? When Can we Consider One Factor to be "Nested" Within Another Factor?

Nested experiments are ones where you may have two or more factors involved and you have a hierarchical order of nesting of factors. This would be different from a factorial design where the factors are varied independently, and the conditions for one factor can be adjusted to be the same at all the other factor levels. In the case where we are doing a variance component analysis of a method take for example the factor "Test Portion." Because each test portion is destroyed in the extraction, we can't really have the exact same test portions for kit Lot 1 as kit Lot 2, so TP will always be a factor nested within some other higher level factor. In the same way, we pipet each extract into 2 wells on the plate to estimate well-to well ELISA variance, since the 2 wells that are used for extract #3 cannot be reused for extract #4, again the factor ELISA is nested within the TP factor. Statisticians will say that for a factor to be nested, there needs to be a significant "separation" in that factor across the different levels of the factor one level higher in the hierarchy. Separation is achieved because the test portion is destroyed and can't be recovered. If a factor is not nested then we say (some authors use this terminology) that the factors are "crossed", meaning they need to be treated as a factorial design, such as a "2x2" factorial. It does not mean to imply there are interactions fitted in the model. To avoid this confusion, some authors refer to these 2 factors as "Main Effects." The area where this will be difficult in these validation designs is the level that includes Analyst/Day/Calibration. For each method and experimental design, we will need to determine if the Analyst/Day factor can be considered nested within the Lot factor, or if there is inadequate separation between Analyst/Days for one lot to another and so will have to be considered as 2 main effects. To make this easy, Lot will

always be a Main Effect at the top of the hierarchy, and TP and ELISA (if replicated) will always be nested. The other easy thing is that the ANOVA calculations in R are simple, and R can do the analysis either way, with a minor change to the code.

Proposed decision rules for determining nested variables are shown in Table D1.

Table D1. Decision rules for determining nested variables

Design	No. analysts	No. days	No. calibrations	Adequate separation?	Factor is
1a or 2a	2	2	1	No	Not nested
1a or 2a	2	2	2	No	Not nested
1a or2a	2	2	4	Yes	Nested
1a or 2a	4	4	2	Yes	Nested
1a or 2a	4	4	4	Yes	Nested
1b or 2b	2	2	1	No	Not nested
1b or 2b	2	2	2	No	Not nested
1b or 2b	2	2	6	Yes	Nested
1b or 2b	4	4	6	Yes	Nested

The idea in Table D1 is if you just have two analysts and 2 days, you can only have enough separation for nested if you have a different calibration for each day/lot combination. If you have four trained analysts in the lab and you can spare them, then you can get separation that way. This is assuming calibration is the significant source of variation, which is usually the case in ELISA methods. In fact, Day is usually always confounded with calibration for a traditional ELISA. The case where there is a common calibration might be if there is a pre-calibrated kit and the calibration is associated with the lot at the factory. If you can't get separation, it is not a problem. You just need to differentiate before the analysis happens so you get the correct ANOVA estimates.

1.4 Model Statements in R

For the nested ANOVA analysis, we will be using R package VCA, which was developed by CLSI for doing method validation on clinical analyses. Information can be found at https://cran.r-project.org/web/packages/VCA/index.html and https://cran.r-project.org/web/packages/VCA/VCA.pdf

(a) *Model statements in R have the general form*.—

Response ~ terms

where "Response" is the numeric response vector and "terms" is a series of terms indicating the predictor variables in some correct syntax dependent on the command being used.

For VCA package in general, we will use the following two types of model statements:

If Analyst is nested: Result ~ Lot/Analyst/TP

If Analyst is not nested: Result ~ (Lot+Analyst)/TP

The names used here such as “Lot,” “Analyst,” “TP” and “Result” are objects defined when the data table is read into the software, and may change depending on the data table.

(b) *Example code for 3-level ANOVA (Designs 1a and 1b).—*

```
Library(VCA)
    Data1<- read.csv("Test Data A1b.csv")
    fit1<- fitVCA(form=Result~(Lot+Analyst), Data=Data1) # Analyst not nested within Lot
    fit2<- fitVCA(form=Result~Lot/Analyst, Data=Data1) # Analyst nested within Lot
```

(c) *Example code for 4-level ANOVA (Designs 2a and 2b).—*

```
    library(VCA)
    Data2<- read.csv("Test Data A2b.csv")
    fit1<- fitVCA(form=Result~(Lot+Analyst)/TP, Data=Data2)   # Analyst not nested within Lot
    fit2<- fitVCA(form=Result~Lot/Analyst/TP, Data=Data2)  # Analyst nested within Lot
```

(d) Note in this package in R (as with most ANOVA procedures in R) you should not include the lowest order factor in the model statement. If you do, the ANOVA table will be incorrect. It is assumed that the lowest factor will be nested. The lowest order factor will be listed in the ANOVA Table as "error.”

1.5 Example Code with Datasets

1.5.1 Data set A1a for Design 1a

See Table D2.

Table D2. Design 1a Data set

Lot	Analyst	TP	Well	Result
1	1	1	1	120.6905
1	1	2	1	108.5775
1	1	3	1	118.6613
1	2	1	1	101.8921
1	2	2	1	106.5847
1	2	3	1	110.5391
2	1	1	1	100.3254
2	1	2	1	109.5876
2	1	3	1	108.2381
2	2	1	1	99.84244
2	2	2	1	95.70943
2	2	3	1	97.8807

Here, Analyst is a stand-in variable name for Analyst/Day/Calibration, TP is the name for Test Portion/extraction. Since there was only 1 well per test portion, the variable "Well" is not really a factor in the experiment.

1.5.2 R-Code for Data Set A1a

```
library(VCA)
DataA1a<- read.csv("Test Data A1a.csv")
fit1<- fitVCA(form=Result~(Lot+Analyst), Data=DataA1a)  # Analyst not nested within Lot
fit1
fit2<- fitVCA(form=Result~Lot/Analyst, Data=DataA1a) # Analyst nested within Lot
fit2.nested
varPlot(form=Result~Lot/Analyst/TP, Data=DataA1a,
    YLabel = list(text="Result", las=0, line=3, cex=1.5),
    Title= list(main="GFA TEST DATA RESULTS PLOT SET A1a", cex.main= 1.75),
    Points= list(pch=20, cex=2.50, col="blue"),
    #MeanLine=list(var="int"),
    MeanLine=list(var=c("Day", "int"), col="blue")
```

1.5.3 Data Output

See Tables D3 and D4.

Table D3. Data output-results: Analyst not nested within Lot

Result ~ (Lot+Analyst)				Analyst not nested within Lot				
	Name	DF[a]	SS[b]	MS[c]	VC[d]	Total, %[e]	SD[f]	CV, %
1	Total	2.68720	95.9797			100	9.79692	9.19518
2	Lot	1	255.407	255.407	39.2024	40.8444	6.26118	5.87661
3	Analyst	1	239.698	239.698	36.5841	38.1165	6.04848	5.67697
4	Error	9	181.739	20.1933	20.1933	21.0391	4.49369	4.21768
Mean	106.5441	(N = 12)						

[a] DF = Degrees of freedom.
[b] SS = Sums of squares.
[c] MS = Mean square error.
[d] VC = Variance component.
[e] Total, % = Percent of total variance contributed by factor.
[f] SD = Standard deviation.

Table D4. Data output-results: Analyst nested within Lot

Result ~ (Lot+Analyst)				Analyst nested within Lot				
	Name	DF[a]	SS[b]	MS[c]	VC[d]	Total, %[e]	SD[f]	CV, %

1	Total	2.955354	77.6877			100	8.81406	8.27269
2	Lot	1	255.408	255.408	22.4709	28.9247	4.74035	4.44919
3	Analyst	2	241.164	120.582	32.6826	42.0693	5.71687	5.36574
4	Error	8	180.273	22.5341	22.534	29.0060	4.74701	4.45544
Mean	106.5441	(N=12)						

[a] DF = Degrees of freedom.
[b] SS = Sums of squares.
[c] MS = Mean square error.
[d] VC = Variance component.
[e] Total, % = Percent of total variance contributed by factor.
[f] SD = Standard deviation.

Use the same code for Design 1b.

1.5.4 Reporting of Precision Estimates

For Designs 1a and 1b, the repeatability standard deviation (s_r) is equivalent to the square root of the test portion variance component, reported as error SD in the ANOVA table.

In the example dataset where analyst is nested within lot, $s_r = 4.75$

For Designs 1a and 1b, the intermediate precision standard deviation (s_i) is equivalent to the square root of the total variance component, reported as total SD in the ANOVA table.

In the example dataset where analyst is nested within lot, $s_i = 8.81$ (*see* Figure D1).

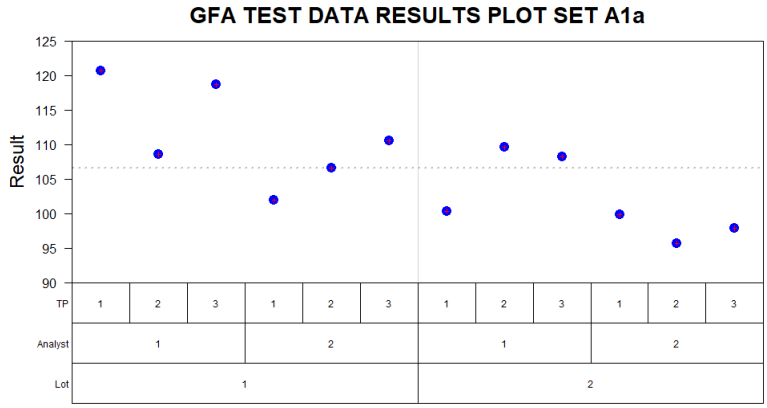


Figure D1. GFA test data results plot set A1a.

1.5.5 Data Set A2b

See Table D5.

Table D5. Data output of data set A2b

Lot	Analyst	TP	Well	Result
1	1	1	1	90.25167
1	1	1	2	89.92019
1	1	2	1	95.44815

1	1	2	2	95.56066
1	2	1	1	84.36506
1	2	1	2	84.57392
1	2	2	1	84.08832
1	2	2	2	84.13355
2	1	1	1	106.9066
2	1	1	2	107.2665
2	1	2	1	109.8504
2	1	2	2	109.1556
2	2	1	1	98.01522
2	2	1	2	98.28006
2	2	2	1	105.577
2	2	2	2	104.6931
3	1	1	1	91.38499
3	1	1	2	94.22005
3	1	2	1	97.7466
3	1	2	2	99.12495
3	2	1	1	92.57129
3	2	1	2	90.96285
3	2	2	1	94.02378
3	2	2	2	94.9194

1.5.6 R-Code for Data Set A2b

```
library(VCA)
DataA2b<- read.csv("Test Data A2b.csv")
fit1<- fitVCA(form=Result~(Lot+Analyst)/TP, Data=DataA2b)  # Analyst not nested
within Lot
fit1
fit2<- fitVCA(form=Result~Lot/Analyst/TP, Data=DataA2b) # Analyst nested within
Lot
fit2
varPlot(form=Result~Lot/Analyst/TP, Data=DataA2b,
      YLabel = list(text="Result", las=0, line=3, cex=1.5),
      Title= list(main="GFA  TEST DATA RESULTS PLOT SET A1b", cex.main= 1.75),
      Points= list(pch=20, cex=2.50, col="blue"),
      #MeanLine=list(var="int"),
      MeanLine=list(var=c("Day",  "int"), col="blue")
)
```

1.5.7 Data Output

See Tables D6 and D7.

Table D6. Data output for Analyst not nested within Lot, TP nested within (Lot+Analyst)

Result ~ (Lot+Analyst)/TP								
	Name	DF[a]	SS[b]	MS[c]	VC[d]	Total, %[e]	SD[f]	CV, %
1	Total	3.182687	92.85136			100	9.635941	10.04162
2	Lot	2	1109.537	554.7684	66.82999	71.97524	8.174961	8.519134
3	Analyst	1	207.8743	207.8743	15.64549	16.85003	3.955437	4.121965
4	Lot:Analyst:TP	8	161.0281	20.12851	9.752615	10.50347	3.122918	3.254395
5	Error	12	7.479314	0.623276	0.623276	0.671262	0.789478	0.822716
Mean	95.96	(N=24)						

[a] DF = Degrees of freedom.
[b] SS = Sums of squares.
[c] MS = Mean square error.
[d] VC = Variance component.
[e] Total, % = Percent of total variance contributed by factor.
[f] SD = Standard deviation.

Table D7. Data output for Analyst nested within Lot and TP Nested within Analyst

Result ~ Lot/Analyst/TP								
	Name	DF[a]	SS[b]	MS[c]	VC[d]	Total, %[e]	SD[f]	CV, %
1	Total	2.958485	85.02862			100	9.221097	9.609313
2	Lot	2	1109.537	554.7684	59.10844	69.51594	7.688202	8.011882
3	Lot:Analyst	3	245.7025	81.90085	15.34189	18.0432	3.916872	4.081776
4	Lot:Analyst:TP	6	123.1998	20.53331	9.955015	11.70784	3.155157	3.287992
5	Error	12	7.479314	0.623276	0.623276	0.733019	0.789478	0.822716
Mean	95.96	(N=24)						

[a] DF = Degrees of freedom.
[b] SS = Sums of squares.
[c] MS = Mean square error.
[d] VC = Variance component.
[e] Total, % = Percent of total variance contributed by factor.
[f] SD = Standard deviation.

Use the same code for Design 2a.

1.5.8 Reporting Precision Estimates

Calculation procedures for repeatability standard deviation and intermediate precision from Designs 2a and 2b depend on whether or not the standard method protocol requires measurement of multiple ELISA wells for each test portion.

For instances where the standard method protocol requires the measurement of multiple replicate ELISA wells (n) for each test portion, with the results averaged to give a single result, the repeatability standard deviation (s_r) is the square root of the sum of the test portion variance component and the ELISA variance divided by the number of replicate wells:

$$s_r^2 = s_{TP}^2 + \frac{s_{ELISA}^2}{n}$$

$$s_r = \sqrt{s_r^2}$$

In the ANOVA table, the test portion variance component is given as the VC for Lot:Analyst:TP. The ELISA variance component is given as the VC for the error row.

For the example dataset when analyst is nested within lot and test portion is nested within analyst:

$$s_r^2 = 9.96 + \frac{0.623}{2}$$

$$s_r = \sqrt{10.27}$$

$$s_r = 3.20$$

For instances where the standard method protocol only requires one ELISA well to be measured for each test portion, the repeatability (s_r) is the square root of the sum of the test portion variance component and the ELISA variance.

$$s_r^2 = s_{TP}^2 + s_{ELISA}^2$$

$$s_r = \sqrt{s_r^2}$$

For the example dataset when analyst is nested within lot and test portion is nested within analyst:

$$s_r^2 = 9.96 + 0.62$$

$$s_r = \sqrt{10.58}$$

$$s_r = 3.25$$

For instances where the standard method protocol requires the measurement of multiple replicate ELISA wells (n) for each test portion, with the results averaged to give a single result, you will want to use designs 2a or 2b, and the intermediate precision standard deviation (s_i) is the square root of the sum of the lot variance component, the analyst variance component, the test portion variance component, and the ELISA variance divided by the number of replicate wells. Do not average the replicate wells before running the ANOVA:

$$s_i^2 = s_{Lot}^2 + s_{Analyst}^2 + s_{TP}^2 + \frac{s_{ELISA}^2}{n}$$

$$s_r = \sqrt{s_r^2}$$

For the example dataset when analyst is nested within lot and test portion is nested within analyst:

$$s_i^2 = 59.11 + 15.34 + 9.96 + \frac{0.623}{2}$$

$$s_r = \sqrt{84.72}$$

$$s_r = 9.20$$

For instances where the standard method protocol only requires one ELISA well to be measured for each test portion (i.e., *n*=1), the intermediate precision standard deviation is equivalent to the square root of the total variance component, reported as total SD in the ANOVA table.

For the example dataset when analyst is nested within lot and test portion is nested within analyst *see* Figure D2.

$$s_r = 9.22$$

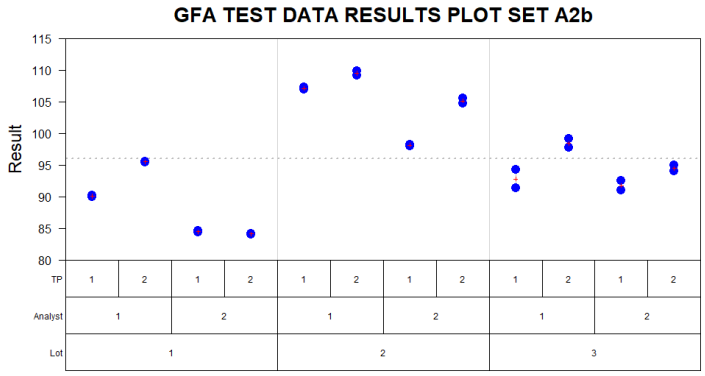


Figure D2. GFA test data results plot set A2b.

2 Limits of Detection and Limits of Quantitation Estimation

Limits of detection (LOD) and quantification (LOQ) should be estimated using methods that account for the relationship between concentration and variance commonly observed with immunoassays (i.e., where variance increases with concentration), as described in IUPAC recommendations (Currie 1999).

2.1 LOD

2.1.1 Data Required

Mean observed concentration and intermediate precision standard deviation from analysis of at least three analyte levels of each claimed matrix, including blank/zero.

2.1.2 Calculations

Calculate mean concentration and intermediate precision standard deviation (S_i) for each test material.

Plot S_i versus observed mean concentration.

Perform linear regression (ordinary least square estimate or weighted least square analysis).

Calculate the LOD according to the following formula (where $\bar{x}_{(0)}$ = calculated mean result from blank samples, $S_{i(0)}$ = intermediate precision standard deviation of blank samples, and slope is the slope from the linear regression above):

$$LOD = \frac{(\bar{x}_{(0)} + 3.3 \times S_{i(0)})}{(1 - 1.65 \times slope)}$$

Values used as estimates of $\bar{x}_{(0)}$ and $S_{i(0)}$ cannot be negative. If the $S_{i(0)}$ from the linear regression (i.e., the intercept value) is negative, use the observed $S_{i(0)}$ from blank matrix samples. If the observed $S_{i(0)}$ also seems unacceptable, use S_i from the lowest concentration test material.

(a) *Multiple matrices.*—Plot S_i versus observed mean concentration for all matrices. It is generally expected that the relationship between S_i and concentration will be sufficiently similar across matrices to conduct a single regression analysis for the combined data from all matrices.

Values used as estimates of $\bar{x}_{(0)}$ and $S_{i(0)}$ cannot be negative. If the $S_{i(0)}$ from the linear regression (i.e., the intercept value) is negative, use the observed $S_{i(0)}$ from blank matrix samples. If the observed $S_{i(0)}$ also seems unacceptable, use S_i from the lowest concentration test material.

For estimation of $\bar{x}_{(0)}$, calculate the mean observed concentration across all blank matrices.

If the relationship between standard deviation and concentration appears to be substantially different between matrices, consult with the AOAC Statistics Committee and Expert Review Panel.

(b) *Example data and calculations*.—*See* Table D8 and Figure D3.

Table D8. Example data table

		0 ppm	0.5 ppm	1.0 ppm	2.5 ppm
Overall mean	$\bar{x}$	0.04	0.612	0.882	2.395
Intermediate precision standard deviation	S_i	0.108	0.211	0.22	0.305
Intermediate precision relative standard deviation	$\%RSD_i$	273.438	34.456	24.888	12.721

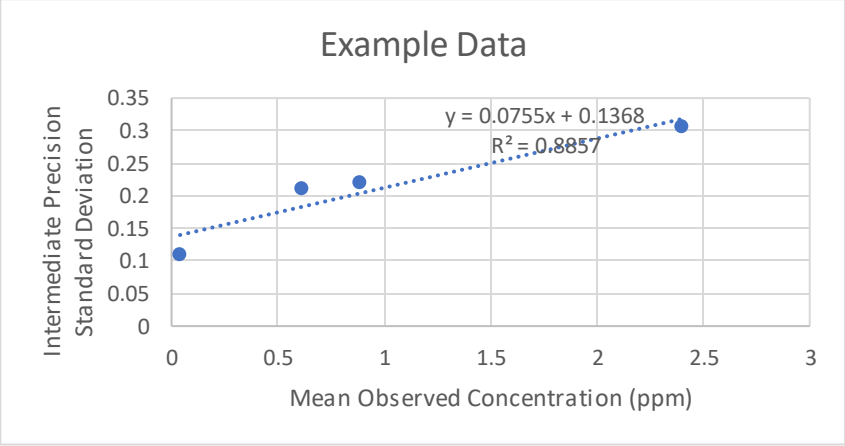


Figure D3. Example data.

$$LOD = \frac{(\bar{x}_{(0)} + 3.3 \times S_{i(0)})}{(1 - 1.65 \times slope)}$$

$$LOD = \frac{(0.04 + 3.3 \times 0.1368)}{(1 - 1.65 \times 0.0755)}$$

$$LOD = 0.56\ ppm$$

2.2 LOQ

2.2.1 Data Required

Mean concentration and intermediate precision standard deviation values from above, linear regression for S_i vs. concentration.

Model %RSD$_i$ across a range of concentrations below and above the expected LOQ using linear regression and $S_{i(0)}$ above.

Linear Regression: $S_i = slope \times concentration + S_{i(0)}$

%RSD$_i$ Modeling: $\%RSD_i = 100 \times \frac{(slope \times concentration + S_{i(0)})}{concentration}$

Plot %RSD$_i$ vs. mean concentration.

Estimate LOQ by calculating the concentration at which %RSD$_i$ would meet acceptable levels (e.g., maximum %RSD$_i$ prescribed in an SMPR). The estimated LOQ must also be greater than the estimated LOD.

Values used as estimates of $S_{i(0)}$ should be the same as those used for the LOD calculations.

$$LOQ = \frac{S_{i(0)}}{\left(\left(\%RSD_i/_{100}\right) - slope\right)}$$

Prepare an operator characteristic (OC) curve for LOQ.—Use a normal distribution calculation function to calculate the probability of obtaining a result higher than the LOQ for the given concentration using the calculated S_i and assuming a normal distribution.

Normal distribution curves can be estimated in Excel with the following function, where LOQ is set at the estimate calculated above, S_i is calculated according to the linear regression:

= 1 – NORMDIST(LOQ, mean concentration, S_i, 1)

Plot the probabilities versus concentration to prepare the OC curve.

2.2.2 Calculations/Procedures for Multiple Matrices

Model %RSD$_i$ across a range of concentrations (as described above) for all matrices together. It is generally expected that the relationship between %RSD$_i$ and concentration will be sufficiently similar across matrices to conduct a single analysis for the combined data from all matrices.

If the relationship between standard deviation and concentration appears to be substantially different between matrices, consult with the AOAC Statistics Committee and Expert Review Panel.

Example data and calculations.—See Table D9 and Figures D4 and D5.

Table D9. Example data

Slope	0.0755
Si(0)	0.1368

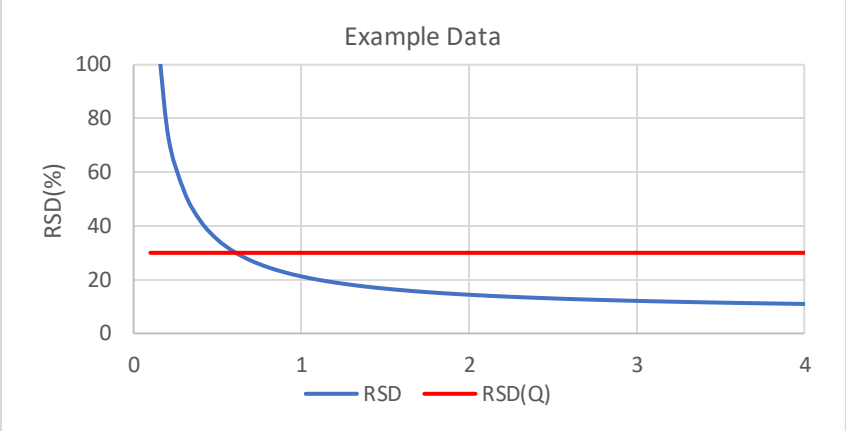


Figure D4. Example data.

$$LOQ = \frac{S_{i(0)}}{\left(\left({}^{\%RSD_i}/_{100}\right) - slope\right)}$$

$$LOQ_{30\%} = \frac{0.1368}{\left(\left({}^{30}/_{100}\right) - 0.0755\right)}$$

$$LOQ_{30\%} = 0.61\ ppm$$

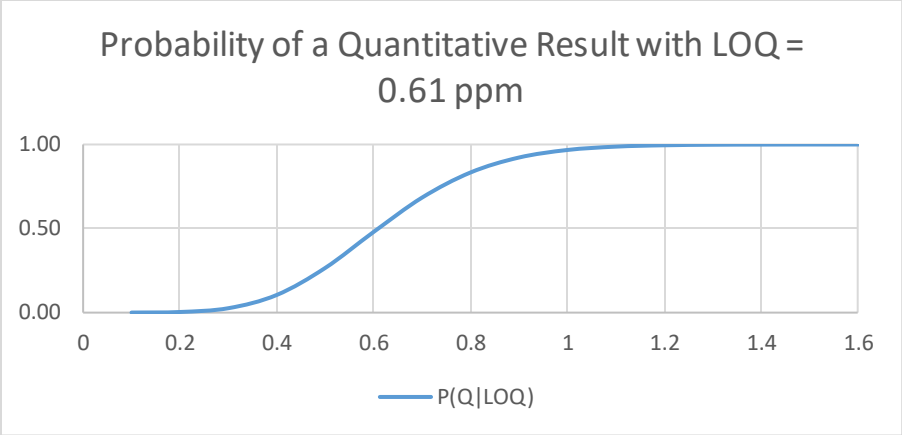


Figure D5. Probability of a quantitative result with LOQ = 061 ppm.

3. Robustness Studies

3.1 Quantitative Assays

For the purpose of the examples provided below, we are using hypothetical assays with written procedures that call for some or all of the following assay parameters, which will be referred to as "factors" going forward: a 2 g sample size, a 30-min extraction, an extraction temperature of 60°C, a diluted sample extract that is loaded onto the ELISA plate in a 100 microliter volume, a 1:10 dilution of the conjugated antibody in conjugate buffer, and/or a 20 min substrate (e.g., TMB) incubation. Remember that each factor should be varied both up and down by at least 20%.

Robustness study designs can be large, and may need to be broken up across test kits, operators, days or other experimental conditions. The factorial treatment combinations (the Runs comprising specific combinations of parameters as shown in the tables below) should be assigned randomly to each set of conditions (test kit, day, operator, etc.).

(a) *Full Factorial (2^y) Designs.*—For assays with just a few factors to vary, a full factorial design (also called a 2^y design) may be used. These designs allow for the analysis of the effects of changes to each individual factor, as well as the analysis of any interactions between factors. Examples of designs where two, three or four factors are varied are given in Tables D10–D12. The row labeled "N" at the bottom of each table describes the "normal" factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

Table D10. Robustness study design example when varying 2 parameters

Run	Sample size, g	Extract time, min
1	1.5	20
2	2.5	20
3	1.5	40
4	2.5	40
N[a]	2	30

[a] N = Normal factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

362

Table D11. Robustness study design example when varying 3 parameters

Run	Sample size, g	Extract time, min	Extract temp., °C
1	1.5	20	45
2	2.5	20	45
3	1.5	40	45
4	2.5	40	45
5	1.5	20	75
6	2.5	20	75
7	1.5	40	75
8	2.5	40	75
N[a]	2	30	60

[a] N = Normal factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

363

Table D12. Robustness study design example when varying 4 parameters

Run	Sample size, g	Extract time, min	Extract temp., °C	Sample load, µL
1	1.5	20	45	50
2	2.5	20	45	50
3	1.5	40	45	50
4	2.5	40	45	50
5	1.5	20	75	50
6	2.5	20	75	50
7	1.5	40	75	50
8	2.5	40	75	50
9	1.5	20	45	150
10	2.5	20	45	150
11	1.5	40	45	150
12	2.5	40	45	150
13	1.5	20	75	150
14	2.5	20	75	150
15	1.5	40	75	150
16	2.5	40	75	150
N[a]	2	30	60	100

[a] N = Normal factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

364

365 **(b)** *Fractional factorial designs.*—When more than four factors are varied, a full-factorial
366 experiment can become prohibitively large. When varying five or more, a fractional factorial

design, or "screening" design, can be used (*see* Tables D13 and D14). A limitation of these designs is that some of the interaction effects may be confounded with each other. But if a resolution IV or V design is used, none of the main effects will be confounded with each other. The following are examples of 2^{5-1} and 2^{6-2} designs, for five and six factors respectively. The 2^{5-1} design is a resolution V, and the 2^{6-2} design is a resolution IV, meaning that main effects are at most confounded with 3rd-order or higher interactions.

Table D13. Varying 5 factors in a 2^{5-1} design, resolution V

Run	Sample size, g	Extract. time, min	Extract temp., °C	Sample load, µL	Conjugate dilution
1	1.5	20	45	50	1:8
2	2.5	20	45	50	1:12
3	1.5	40	45	50	1:12
4	2.5	40	45	50	1:8
5	1.5	20	75	50	1:12
6	2.5	20	75	50	1:8
7	1.5	40	75	50	1:8
8	2.5	40	75	50	1:12
9	1.5	20	45	150	1:12
10	2.5	20	45	150	1:8
11	1.5	40	45	150	1:8
12	2.5	40	45	150	1:12
13	1.5	20	75	150	1:8
14	2.5	20	75	150	1:12
15	1.5	40	75	150	1:12
16	2.5	40	75	150	1:8
N[a]	2	30	60	100	1:10

[a] N = Normal factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

Table D14. Varying 6 factors in a 2^{6-2} design, resolution IV

Run	Sample size, g	Extract time, min	Extract temp., °C	Sample load, µL	Conjugate dilution	TMB time, min
1	1.5	20	45	50	1:8	15
2	2.5	20	45	50	1:12	25
3	1.5	40	45	50	1:12	25
4	2.5	40	45	50	1:8	15
5	1.5	20	75	50	1:12	15
6	2.5	20	75	50	1:8	25
7	1.5	40	75	50	1:8	25
8	2.5	40	75	50	1:12	15
9	1.5	20	45	150	1:12	25
10	2.5	20	45	150	1:8	15
11	1.5	40	45	150	1:8	15
12	2.5	40	45	150	1:12	25
13	1.5	20	75	150	1:8	25
14	2.5	20	75	150	1:12	15
15	1.5	40	75	150	1:12	15
16	2.5	40	75	150	1:8	25
N[a]	2	30	60	100	1:10	20

[a] N = Normal factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

Additional fractional factorial designs can be generated in R – see the instructions later in this document.

(c) *Plackett-Burman designs.*—When more than six factors are varied, a Plackett-Burman design may be needed. This type of design reduces the total number of experimental runs while still allowing for the analysis of the main effects of individual factors. A concern with Plackett-Burman designs is that the main effects of the individual factors are confounded with interaction effects, so it can't be determined if any significant effects are due to changes in an individual factor, or to that plus the changes in another factor. If it is important to discriminate between the individual effects and interaction effects, then the few significant factors identified by the Plackett-Burman design may be used in a separate full factorial experiment.

Examples of designs where seven or eight factors are varied are given in Tables D15 and D16 (our hypothetical method only had six factors to vary, so factors seven and eight are unnamed). With this same 12-row design you may test up to 11 factors – examples of this can be found in the NIST Engineering Statistics Handbook (https://web.archive.org/web/20220923135605/https:/www.itl.nist.gov/div898/handbook/pri/section3/pri335.htm). The row labeled "N" at the bottom of each table describes the "normal" parameter values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

Table D15. Varying 7 parameters in a Plackett-Burman design

Run	Sample size, g	Extract. time, min	Extract temp., °C	Sample load, µL	Conjugate dilution	TMB time, min	Factor 7
1	2.5	40	75	150	1:8	25	High
2	1.5	40	45	150	1:8	25	Low
3	1.5	20	75	50	1:8	25	High
4	2.5	20	45	150	1:12	25	High
5	1.5	40	45	50	1:8	15	High
6	1.5	20	75	50	1:12	25	Low
7	1.5	20	45	150	1:12	15	High
8	2.5	20	45	50	1:8	15	Low
9	2.5	40	45	50	1:12	25	Low
10	2.5	40	75	50	1:12	15	High
11	1.5	40	75	150	1:12	15	Low
12	2.5	20	75	150	1:12	15	Low
N[a]	2	30	60	100	1:10	20	Mid

[a] N = Normal factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

Table D16. Varying 8 parameters in a Plackett-Burman design

Run	Sample size, g	Extract. time, min	Extract temp., °C	Sample load, µL	Conjugate dilution	TMB time, min	Factor 7	Factor 8
1	2.5	40	75	150	1:8	25	High	High
2	1.5	40	45	15	1:8	25	Low	Low
3	1.5	20	75	50	1:8	25	High	Low
4	2.5	20	45	150	1:12	25	High	High
5	1.5	40	45	50	1:8	15	High	High
6	1.5	20	75	50	1:12	25	Low	High
7	1.5	20	45	150	1:12	15	High	Low
8	2.5	20	45	50	1:8	15	Low	High
9	2.5	40	45	50	1:12	25	Low	Low
10	2.5	40	75	50	1:12	15	High	Low
11	1.5	40	75	150	1:12	15	Low	High
12	2.5	20	75	150	1:12	15	Low	Low
N[a]	2	30	60	100	1:10	20	Mid	Mid

[a] N = Normal factor values for the assay, but this is only included for informational purposes and does not need to be tested as part of the experiment.

3.2 Result Reporting

All results from the robustness study should be reported. Table D17 is an example result table. This is a result table from a full factorial design that varied 3 parameters, with 5 replicates per factorial pattern, but a similar table design can be used for any size full-factorial or Plackett-Burman design.

Table D17. Robustness study **e**xample result table

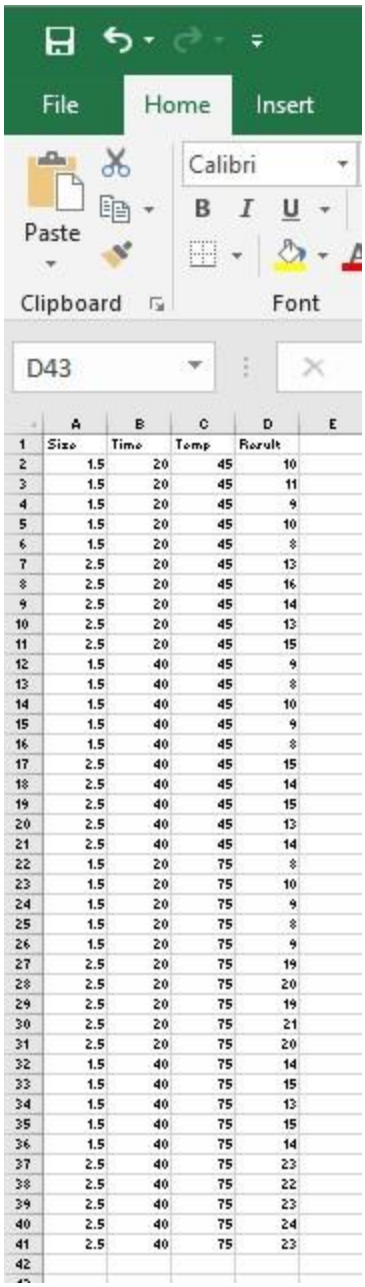
Run	Sample size, g	Extract time, min	Extract temp., °C	Test portion results – 20 ppm sample				
				1	2	3	4	5
1	1.5	20	45	10	11	9	10	8
2	2.5	20	45	13	15	14	13	15
3	1.5	40	45	9	8	10	9	8
4	2.5	40	45	15	14	15	13	14
5	1.5	20	75	8	10	9	8	9
6	2.5	20	75	19	20	19	21	20
7	1.5	40	75	14	15	13	15	14
8	2.5	40	75	23	22	23	24	23

3.3 Study Analysis

Both the factorial and Plackett-Burman designs can be analyzed using a linear regression or factorial ANOVA.

The following is an example of performing a factorial ANOVA on the data set in Table D17, using R and RStudio. R and RStudio are free, open access programs that can be used online at https://www.rstudio.com/products/cloud/, or downloaded from https://posit.co/download/rstudio-desktop/. Because RStudio works by writing lines of code, it is helpful to shorten the names of the column headers.

In order to run the ANOVA, the data from Table D17 needs to be reoriented into a “long” format, with each test result on its own individual row. When you enter the factor levels, remove any lettering or special characters, and just enter the numbers. Figure D6 is an example of how this might appear on an Excel sheet, but RStudio also allows you to upload data sets from text files and other statistics programs, or to enter the data in manually (you can learn more about using RStudio at https://education.rstudio.com/learn/beginner/). However you save your dataset, you want to give it a short file name as well. In this example, we are naming the Excel file “data1”.



	A	B	C	D	E
1	Size	Time	Temp	Result	
2	1.5	20	45	10	
3	1.5	20	45	11	
4	1.5	20	45	9	
5	1.5	20	45	10	
6	1.5	20	45	8	
7	2.5	20	45	13	
8	2.5	20	45	16	
9	2.5	20	45	14	
10	2.5	20	45	13	
11	2.5	20	45	15	
12	1.5	40	45	9	
13	1.5	40	45	8	
14	1.5	40	45	10	
15	1.5	40	45	9	
16	1.5	40	45	8	
17	2.5	40	45	15	
18	2.5	40	45	14	
19	2.5	40	45	15	
20	2.5	40	45	13	
21	2.5	40	45	14	
22	1.5	20	75	8	
23	1.5	20	75	10	
24	1.5	20	75	9	
25	1.5	20	75	8	
26	1.5	20	75	9	
27	2.5	20	75	19	
28	2.5	20	75	20	
29	2.5	20	75	19	
30	2.5	20	75	21	
31	2.5	20	75	20	
32	1.5	40	75	14	
33	1.5	40	75	15	
34	1.5	40	75	13	
35	1.5	40	75	15	
36	1.5	40	75	14	
37	2.5	40	75	23	
38	2.5	40	75	22	
39	2.5	40	75	23	
40	2.5	40	75	24	
41	2.5	40	75	23	
42					

Figure D6. 3-factor study from Table 8, with data in the "long" format required for analysis in R

Once you have both R and Rstudio downloaded and installed, open RStudio and find the Environment tab. Click on Import Data Set, then From Excel, and follow the instructions to import your file. In the Environment window you should then see your data1 dataset, and if you click the drop-down arrow to the left it will show you a data summary as shown in Figure D7.

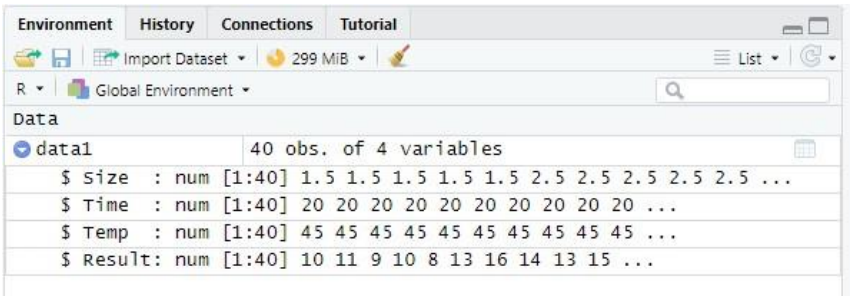


Figure D7. Data summary view in R, of data from Figure 1

The following is a list of the commands you will enter to perform the analysis of variance. Hit Enter after typing in each command.

>library(car)

>Anova(lm(Result ~ Size + Time + Temp, data = data1)

[in this line of code, "lm" is telling it to run a linear model; "Result" is your result column, the dependent variable; "Size", "Time", and "Temp" are the column titles for your independent variables, and "data1" is the name of your dataset]

The results of the analysis are displayed as shown in Table D18 below:

Table D18. Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)[a]
Size	555.02	1	133.117	1.191e-13 ***
Time	38.02	1	9.1199	0.004629 **
Temp	225.63	1	54.1139	1.126e-08 ***
Residuals				

---

[a] Signif. codes = 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

3

The p-values for each factor are shown in the Pr(>F) column. Significance is determined as $p<.05$, but you have done three comparisons here, so you will want to make a Bonferroni adjustment, and consider a factor to be significant when the p-value is less than 0.017 (i.e., 0.05 divided by 3). In this example, changes to each of the factors makes a significant impact on the results. This means that the method instructions should warn the end user to avoid deviations in any of these steps.

If you were analyzing more than the three factors, the code and commands would be the same, you would simply enter the additional column titles after the ~ sign using + signs in between.

If you were interested in seeing the interaction effects between each of the factors, you would use the following code:

>Anova(lm(Result ~ Size*Time*Temp, data = data1)

3.4 Generating Fractional Factorial Designs in R

You can generate additional fractional factorial designs in R using the following commands. The example here is for a 2^{5-1} design, which has a level V resolution – this is the same example shown in Table D13:

>install.packages("FrF2")

>library(FrF2)

>fivefactors <- FrF2(nfactors = 5, resolution = 5, randomize = FALSE)

[In this code, "fivefactors" is a name you make up to describe the table we are trying to generate, nfactors is the number of factors you are varying, and resolution is your chosen resolution (you will want to keep it at 4 or 5 to avoid having main effects confounded with each other).]

>summary(fivefactors)

[the summary command gives you the results of the analysis, shown below]

Call:

FrF2(nfactors = 5, resolution = 5, randomize = FALSE)

Experimental design of type FrF2

16 runs

Factor settings (scale ends):

	A	B	C	D	E
1	–1	–1	–1	–1	–1
2	1	1	1	1	1

Design generating information:

$legend

[1] A=A B=B C=C D=D E=E

$generators

[1] E=ABCD

Alias structure:

[[1]]

[1] no aliasing among main effects and 2fis

The design itself is show in Table D19.

Table D19. Example study design export from R for a 2^{5-1} design with level V resolution

	A	B	C	D	E
1	–1	–1	–1	–1	1
2	1	–1	–1	–1	–1
3	–1	1	–1	–1	–1
4	1	1	–1	–1	1
5	–1	–1	1	–1	–1
6	1	–1	1	–1	1
7	–1	1	1	–1	1
8	1	1	1	–1	–1
9	–1	–1	–1	1	–1
10	1	–1	–1	1	1
11	–1	1	–1	1	1
12	1	1	–1	1	–1
13	–1	–1	1	1	1
14	1	–1	1	1	–1
15	–1	1	1	1	–1
16	1	1	1	1	1

[a]*Note*: In this result display, –1 indicates the lower level of the factor, and 1 indicates the higher level of the factor. Compare to Table D13.

class=design, type= FrF2.