TECHNICAL SPECIFICATION

ISO/TS 16393

First edition
2019-02

Molecular biomarker analysis — Determination of the performance characteristics of qualitative measurement methods and validation of methods

Analyse de biomarqueurs moléculaires — Détermination des caractéristiques de performance des méthodes de mesure qualitatives et validation des méthodes

Reference number
ISO/TS 16393:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 16, *Horizontal methods for molecular biomarker analysis*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Qualitative (binary) analytical methods (e.g. applied to screening tests) for use in the analysis of food or food products (including seeds of food crops) with the purpose of demonstrating the presence/absence of a given measurand in a sample should provide objective evidence that they are adequate for their intended use. A validated test method is much preferred over one that has not undergone studies to determine its accuracy and reliability for its specific purpose. These methods that yield a binary result (yes/no, positive/negative, etc.) are referred to as “qualitative” or “binary” methods.

As with quantitative methods, qualitative method performance has to be characterized with respect to the concentration of the measurand. However, only two conditions are indicated in the result: either the measurand is detected (a positive result) or it is not detected (a negative result). While internationally recognized guidelines (e.g. ISO 5725-2, References [7] and [16]) have been produced over the years to harmonize the validation of quantitative analytical methods, no consensus is yet available among stakeholders on a practical implementation of the performance criteria approach to the validation of qualitative methods for use in food and food products.

Conceptual approaches for validating qualitative methods classically focused on parameters such as sensitivity, selectivity, false positive rate and false negative rate, based on detection/non-detection of the measurand in the test sample. The limitation of this approach was the underlying assumption that the method had a predictable response to the presence of a measurand present at a non-zero concentration. In practice, however, a non-zero concentration can result in a variable probability of a positive result in the assay. Treating the concentration of measurand as a continuous variable with reasonable and/or previously determined confidence in a defined matrix using a specific analytical method is a better predictor of measurement response than a two-state, zero/non-zero variable.

This document describes the assessment of probability of detection (POD). This approach allows for comparison of probabilities across concentrations and further allows for a simple graphical representation of validation data as a POD response curve graphed by concentration with associated error bars of the mean POD value. This approach expresses the POD as dependent on concentration; the goal of validation is to characterize the response probability curve as a function of measurand mass or concentration.

A number of models have been described in the literature for the calculations of the confidence intervals of the POD and confidence intervals or predictive ranges for concentrations in case of a positive or negative result, e.g. References [4], [8], [9], [11], [17], [19] and [20]. Whereas qualitative methods are often evaluated at 50 %, they are used close to 100 %, or at levels where the sample size is adjusted so as to always obtain a clear positive or negative result. The present specification is therefore the result of an extensive discussion of the possible improved models for characterization of qualitative methods, particularly focused on the characterization of the methods close to the 0 and 100 % POD cases. The performance characteristics include:

a) the mean POD across laboratories (LPOD);

b) the corresponding confidence interval of the LPOD, which is the interval estimate of the mean POD;

c) the prediction interval for future observations of laboratory specific PODs.

An advanced statistical method allows the user to calculate confidence and/or prediction intervals for the concentrations where the user would expect positive or negative results. To do so is particularly challenging where the POD is close to 0 % or 100 %.

Molecular biomarker analysis — Determination of the performance characteristics of qualitative measurement methods and validation of methods

1 Scope

This document specifies methods that yield a binary result and are used for the determination in food or food products (including seeds of food crops) of the presence of molecular biomarkers. These methods are typically used where the measurand is expected to be present in very small amounts and concentrations at the limit of detection (LOD).

Methods are validated in terms of the probability of detection (POD) and of the precision of the POD. They do not rely on the concept of false positive/false-negative results, or the concept of LOD. However, inferences about the precision of the classical LOD can be made.

This document describes the extent of method validation. The annexes provide different statistical models that can be considered depending on the analytical method, structure of data and statistical experience.

This document does not apply to quantitative methods that are used to make a detection decision by comparing the value of a response to a cut-off value using a quantitative method, where the methods are validated by using quantitative statistics on the responses. This document also does not apply to microbiological test methods, starch, essential oils or quantitative methods.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 5725-1:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO 5725-2:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
binary result
result from a *method* (3.6) of analysis where there are only two possible outcomes

3.2
intraclass correlation coefficient
ICC
measure of the reliability of measurements (between laboratories)

Note 1 to entry: The coefficient represents agreements between two or more results measured on identical samples.

3.3
identical test item
sample that is prepared and can be presumed to be identical for the intended purpose of measurement of the measurand (and can be presumed to be identical for the intended purpose)

[SOURCE: ISO 3534-2:2006, 1.2.34, modified — "and can be presumed to be identical purpose of measurement of the measurand" has been added and Note 1 to entry has been deleted.]

3.4
lower confidence limit
LCL
$\hat{\mu}_L$
lower value of a range containing the true value of the measurand with a specified probability

Note 1 to entry: The symbol for LCL is taken from Reference [5].

3.5
mean probability of detection across laboratories
LPOD
$P_{\alpha\lambda}$
probability of a positive analytical outcome of a *qualitative method* (3.9) for a given matrix at a given concentration in multiple laboratories

Note 1 to entry: Throughout this document, when used in mathematical formulae, $P_{\alpha\lambda}$ refers to the estimator for the *probability of detection (POD)* (3.8) parameter across laboratories.

Note 2 to entry: The symbol for LPOD is the symbol for POD with the lowercase Greek letter ƛ (lambda) to indicate laboratory-wide.

3.6
method
procedure that includes sample processing, assay and data interpretation

3.7
naturally incurred sample
sample that contains the measurand by virtue of its inherent characteristics rather than the measurand being intentionally added

3.8
probability of detection
POD
P_{α}
probability of a positive analytical outcome of a *qualitative method* (3.9) for a given matrix at a given concentration in a single laboratory

Note 1 to entry: Throughout this document, when used in mathematical formulae, P_{α} refers to the estimator for the probability of detection parameter.

Note 2 to entry: The symbol for POD is drawn from the term P for probability and the first letter of the Greek term for detection, ανίχνευση.

3.9
qualitative method
method (3.6) of analysis with two possible outcomes

Note 1 to entry: Qualitative method is an alternative terminology to binary method.

3.10
replicate test sample
sample taken from a bulk sample such that the replicate test samples are as close to identical as achievable, in order to constitute *identical test items* (3.3)

3.11
validation experiment
determination of *method* (3.6) performance parameters from a series of test results reported by one or more usually a number of participating laboratories

3.12
upper confidence limit
UCL
$\hat{\mu}_U$
upper value of a range containing the true value of the measurand with a specified probability

Note 1 to entry: The symbol for UCL is taken from Reference [5].

4 Characterization of a qualitative method via a validation experiment

4.1 Criteria for a standard measurement method

The following criteria should be taken into consideration when validating a qualitative method of analysis:

— applicability;

— robustness;

— selectivity;

— POD related to the measurand concentration.

All measurements shall be carried out according to a standard method based on a written document that describes in full detail how the measurement shall be carried out, including the applicability and selectivity of the method. It shall incorporate information based on the robustness testing of the method established at the single laboratory level when developing the method. The standard method may be modified by the result of experiments to determine the intermediate precision and/or the results of collaborative multi-laboratory trial(s).

4.2 Performance of a validation experiment

The estimates of performance parameters derived from a validation experiment are valid only for tests carried out according to the standard measurement method. A validation experiment can be considered to be a practical test of the adequacy of the standard measurement method. One of the main purposes of standardization is to standardize how methods are characterized, and eliminate differences between users (laboratories) as far as possible. The data provided by a validation experiment will reveal how effectively this purpose has been achieved. Pronounced differences between the laboratories often indicate that the measurement method can be improved.

From a practical point of view, it is important and desirable to carry out a number of steps before proceeding with the validation experiment. This includes: a) measurement of several replicates by one operator to establish suitable test materials that will cover the desired POD levels, followed by: b) a

mini validation experiment to establish that the instructions for the experiment are clear and sufficient and that the test materials are suitable for the full validation experiment.

4.3 Nature of test materials

Validation of qualitative methods requires the use of known positive (low and high POD) and negative (effectively as close as possible to zero POD) materials. Special challenges arise when a biological material is being tested, and pure reference material (CRM traceable back to SI units) may not be readily available. For some biomolecular methods, naturally incurred samples may be the only source of materials for validation. The preparation and source of each material shall be documented. Wherever possible, a quantitative method can be used to confirm the concentration of the measurand.

4.4 Requirements for replicate test samples

In a validation experiment, a number of replicate test samples of a specific material or specimens of a specific product are typically sent from a central point to a number of laboratories. The definition of repeatability conditions states that the measurements in these laboratories shall be performed on identical test items and refers to the moment when these measurements are actually carried out.

The test materials will ideally be evaluated for homogeneity before preparing the replicate laboratory samples to be sent to the laboratories, or by testing a number of the replicate test samples if a suitable method is available. Furthermore, the replicate test samples shall be identical test items (under the definition of ISO 5725-1) when dispatched to the laboratories and the replicate samples shall be stable and remain identical during transport and during the different time intervals that can elapse before the measurements are actually performed.

NOTE 1 The terms “identical” and “identical test items” are not the same as “identical test portions” (see ISO 5725-2:1994, Clause 5). There will always be some level of variation between replicate test samples (i.e. the actual materials sent), and this is an integral part of testing method repeatability. Test portion variability is dependent on concentration, test portion size and matrix homogeneity. When preparing the replicate test samples for a collaborative study, the concept of identical test items is to be interpreted as each test sample having an equal probability of producing a positive test result. This means that all laboratories receive essentially the same test items. The test portions will always have some level of variation, which is an inherent part of the measurement variation.

NOTE 2 The number of replicate samples required to get a good estimation (at 95 % confidence) of the LPOD for a two-sided coverage is 12 per level for the range 25 % to 75 % LPOD for the case where 8 laboratories are included (see Table E.2). If more participants are available, the number of replicate samples can be lowered in consultation with a statistician. However, the larger numbers needed to get ideal estimates of the LPOD at high and low measurand concentrations may not be practicable to achieve in a multi-laboratory trial.

Conditions should be representative of the use of the method in the laboratory. It shall be clearly stated when reporting the results if an intermediate material, such as a ground sample or an extract, is distributed for this purpose. Moreover, it shall be shown that the intermediate materials are stable under shipping conditions.

NOTE 3 While the replicate test samples supplied at each concentration would preferably consist of unprocessed material (such as whole grain or seeds) in order to test the whole method from sample to result, this is, in most cases, impractical. Therefore, it is most practical to grind the material and distribute a typical powder that would be obtained under typical conditions.

Test materials are prepared and divided into test samples before these replicate test samples are shipped to the participating laboratories. The replicate test samples may be reduced to test portions in the laboratory or analysed directly. The relationship is given in Figure 1.

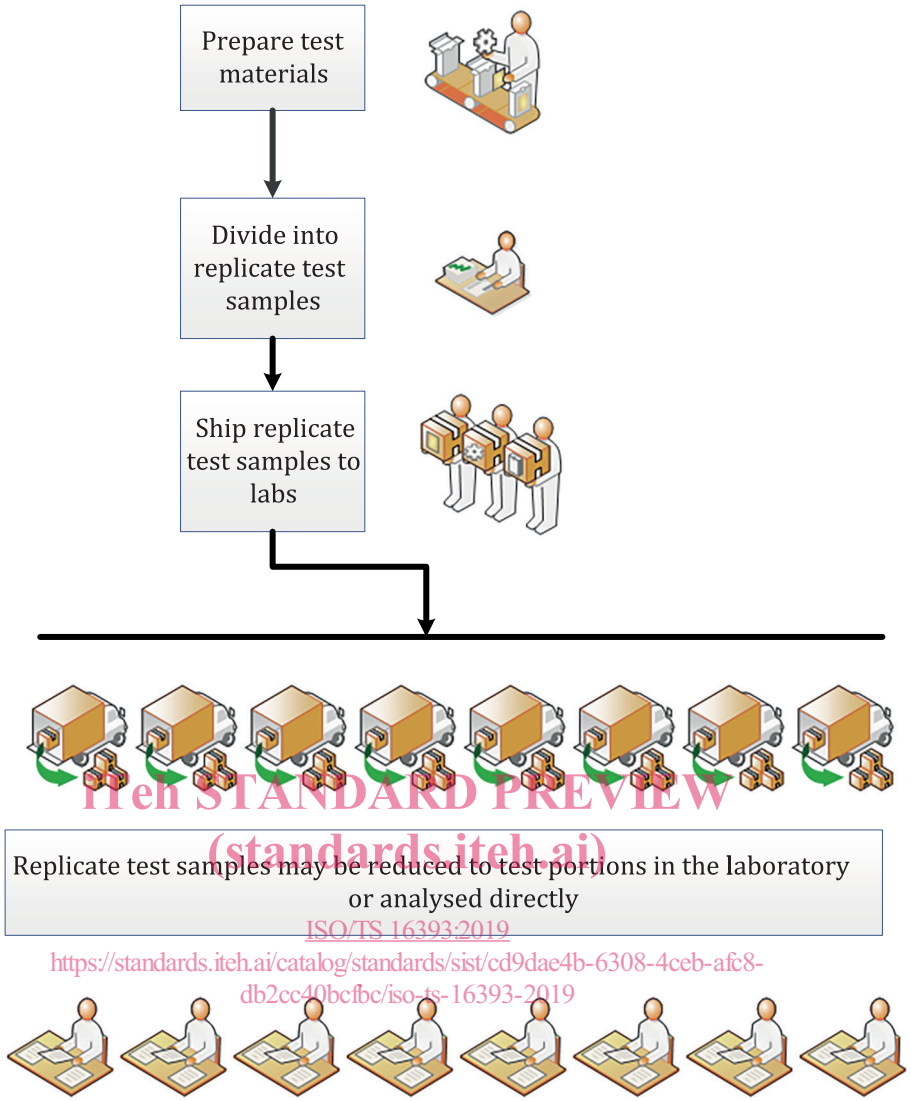

Figure 1 — Relationship between the test materials, replicate test samples and test portions

4.5 Robustness (ruggedness)

The method developer is expected to evaluate the robustness of the method against small changes in analytical conditions and external influences, and identify variables which could have a significant effect on method performance. Critical variables should be included in the standard measurement method (e.g. by including an acceptable temperature range).

4.6 Applicability

The user should be able to determine whether the method will be appropriate for the desired application (fit for purpose) and if there will be limitations to its use. Applicability is the analytes, matrices and concentrations for which a method of analysis may be used satisfactorily. An applicability statement shall therefore be provided by the method developer. It should include a list of the known analyte(s) or measurand(s) that can be determined by the method, and the form in which analyte(s) can be determined, e.g. speciation, total/available, the sample matrix(es) within which those analyte(s) can be determined. In addition to a statement of the range of capability of satisfactory performance for each factor, the statement of applicability may also include warnings as to known interference by other analytes, or inapplicability to certain matrices and situations. For example, concentrations that may lead to reduced POD at concentrations higher than those normally expected should also be specified, as

certain methods (such as those depending on antibodies) have the possibility of giving a negative result at very high concentrations of the measurand (the hook effect).

NOTE Applicability outside of the food sector can be referred to as “scope”.

4.7 Selectivity

Determination of selectivity is a single laboratory study designed to demonstrate that a method does not detect non-target measurands expected to erroneously give a positive result due to chemical or structural similarities.

The method should be shown to give a positive result for claimed measurands. Each measurand from the selectivity test panel should be tested at the appropriate target concentration for each measurand.

4.8 Experimental design for a multi-laboratory study

4.8.1 Participating laboratories

Ideally, the chosen laboratories should be a random sampling of all potential method users. Laboratories participating in any validation study for qualitative methods should have experience and training in performing the type of method being tested. However, the participating laboratories should not consist exclusively of those that have gained special experience during the process of standardizing the method. Neither should they consist (exclusively) of specialist reference laboratories, in order to demonstrate the accuracy to which the method can perform in expert hands.

Estimating the POD at applicable measurand concentrations can be carried out provided that an adequate number of replicate test samples are analysed across a suitable number of concentrations and a sufficient number of laboratories. The number of replicates per laboratory and the number of laboratories should be chosen with consideration of the effect of the size of the validation experiment on the size of the confidence intervals that will be obtained.

4.8.2 Number of laboratories

The purpose of involving a large number of laboratories in the study is to get a wider subset of potential method users to contribute data to the study. Using a large number of laboratories will reduce the subsampling error and will mean that the estimates that are obtained in the study will be less biased. In addition, with more laboratories, it will be easier to detect a laboratory effect in the data, if it is significant. The absolute minimum number of laboratories reporting and included in the final statistical analysis of the study is eight.

4.8.3 Number of levels

The minimum number of concentration levels to study should be five.

The experiment should verify that the method is sensitive to concentration, so that at low levels there is a low POD and that at a high concentration there is a POD. The experiment shall be designed to best characterize the POD curve, in as efficient a manner as possible.

One concentration level should be chosen where the expected POD is close to zero. This will demonstrate the method will not give a positive response at low, near-zero concentrations.

There should be a second concentration level where the method is expected to give > 95 % of positive responses.

There will be some concentration levels where the POD is expected to be in a marginal range (0,85 to 0,95 or 0,05 to 0,15), which is important to identify so that the response curve can be better characterized and the transition concentration from medium POD to high POD can be identified. In addition, a sample in the mid-range (35 % POD to 65 % POD) will allow the experiment to expose cases where there is a large difference in sensitivity between participating laboratories.

