GUIDE 35

Reference materials — General and statistical principles for certification

Third edition 2006

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword ... v
Introduction ... vi
1 Scope ... 1
2 Normative references ... 1
3 Terms, definitions and symbols ... 2
4 Symbols ... 3
5 Design of a certification project ... 4
5.1 General ... 4
5.2 Project definition ... 5
5.3 Transport issues ... 5
5.4 Collection of starting material ... 5
5.5 Feasibility study ... 6
5.6 Required lifetime and shelf life ... 6
5.7 Sample preparation ... 6
5.8 Homogeneity study ... 8
5.9 Stability study ... 8
5.10 Choice of measurement methods ... 9
5.11 Certification ... 10
5.12 Summary of project design ... 10
6 Evaluating measurement uncertainty ... 11
6.1 Basis for evaluating the uncertainty of a property value of a (C)RM ... 11
6.2 Basic model for a batch characterization ... 12
6.3 Uncertainty sources ... 13
6.4 Issues with distribution functions ... 13
6.5 Use of ratios ... 14
6.6 Choice of a coverage factor ... 14
6.7 Recertification ... 14
7 Homogeneity study ... 15
7.1 Preamble ... 15
7.2 Materials ... 15
7.3 Concept of homogeneity ... 16
7.4 Practice ... 16
7.5 Measurements ... 16
7.6 Statistically valid sampling schemes and trend analysis ... 17
7.7 Evaluating a homogeneity study ... 17
7.8 Between-bottle homogeneity study ... 18
7.9 Insufficient repeatability of the measurement method ... 19
7.10 Within-bottle homogeneity ... 19
8 Stability study ... 20
8.1 Types of (in)stability ... 20
8.2 Designs of experiments ... 21
8.3 Evaluation of results ... 22
8.4 Stability monitoring ... 25
8.5 Determination of the shelf life in relation to the long-term stability ... 27
9 Determination of the property values ... 27
9.1 General ... 27
9.2 Establishing and demonstrating traceability ... 28

9.3 Practical approaches **30**
9.4 Measurement design **31**
9.5 Property-related considerations **35**

10 Data and uncertainty evaluation **38**
10.1 Models **38**
10.2 Data formats **38**
10.3 Distributions **40**
10.4 Data screening **41**
10.5 Data evaluation **41**
10.6 Uncertainty evaluation **43**
10.7 Uncertainty-based evaluation **44**
10.8 Specific issues **45**

11 Certification **46**

Annex A (informative) **Statistical approaches** **48**

Annex B (informative) **Examples** **53**

Bibliography **62**

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

Draft Guides adopted by the responsible Committee or Group are circulated to the member bodies for voting. Publication as a Guide requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO Guide 35 was prepared by the ISO *Reference Materials Committee* (REMCO).

This third edition cancels and replaces the second edition (ISO Guide 35:1989), of which all clauses referring to the estimation of measurement uncertainty have been thoroughly revised. This revision also provides an up-to-date description of the technical issues related to the production and certification of reference materials.

Introduction

The production, characterization and certification of reference materials (RMs) is a key activity in improving and maintaining a worldwide coherent system of measurements. As detailed in ISO Guide 32 and ISO Guide 33, certified reference materials (CRMs) are used for calibration, quality control and method validation purposes, as well as for the assignment of values to other materials, which in turn can also be CRMs. Furthermore, CRMs are used to maintain or establish traceability to conventional scales, such as the octane number, hardness scales and pH. Last, but not least, selected pure substances are also used to maintain the international temperature scale.

For producers of CRMs, there are three ISO Guides that assist the set-up of a facility to produce and certify RMs and to ensure that the quality of thus-produced CRMs meet the requirements of the end-users. ISO Guide 34 outlines the requirements to be met by a CRM producer to demonstrate competence, whereas this Guide provides assistance on how to meet these requirements. At a fairly generic level, this Guide provides models for homogeneity testing, stability testing, and the characterization of the candidate CRM. ISO Guide 31 describes the format and contents of certificates for CRMs.

In some ways, this Guide can be seen as an application of the *Guide to the Expression of Uncertainty in Measurement* (GUM) with respect to the peculiarities of the production of CRMs. Where possible, this Guide makes reference to the GUM, as the latter describes in detail how to evaluate measurement uncertainty of a value obtained from measurement. This Guide complements the GUM in a sense that it provides additional guidance with respect to the inclusion of the uncertainties due to the (remaining) batch inhomogeneity and instability of the CRM in the uncertainty of the property values, and the determination of these uncertainty contributions.

Although this Guide has been developed to support best practice in the production and characterization of RMs, using it without carefully considering whether specific parts are applicable to the particular CRM may still cause its property values (and their uncertainties) to be established on a wrong or faulty basis. A user of this type of documentation should consider that it cannot substitute for “critical thinking, intellectual honesty and professional skill” (GUM:1993, 3.4.8). The quality of the “product” CRM depends as much on these aspects as on the use of proper procedures and methods.

Thorough knowledge of the material and its properties, and of the measurement methods used during homogeneity testing, stability testing and characterization of the material, along with a thorough knowledge of the statistical methods, are needed for correct processing and interpretation of experimental data in a typical certification project. It is the combination of these required skills that makes the production and certification of RMs so complex. The greatest challenge in these projects is to combine these skills to allow a smooth implementation of the project plan.

Most of the contents of this Guide can be applicable to the production of RMs. Requirements such as the traceability of the property values, the necessity of a full evaluation of measurement uncertainty, among others, apply to most categories of RMs to serve, for example, as calibrants or as a means to check the performance of a method, or to assign a value to another material.

Pharmacopoeial standards and substances are established and distributed by pharmacopoeial authorities following the general principles of this Guide. Specific guidance for the production of these kinds of RMs exists. It should be noted, however, that a different approach is used by the pharmacopoeial authorities to give the user the information provided by certificates of analysis and expiration dates. Also, the uncertainty of their assigned values is not stated since it is not permitted by the prescribed use of these RMs in the relevant compendia.

Reference materials — General and statistical principles for certification

1 Scope

This Guide gives statistical principles to assist in the understanding and development of valid methods to assign values to properties of a reference material, including the evaluation of their associated uncertainty, and establish their metrological traceability. Reference materials (RMs) that undergo all steps described in this Guide are usually accompanied by a certificate and called a certified reference material (CRM). This Guide will be useful in establishing the full potential of CRMs as aids to ensure the comparability, accuracy and compatibility of measurement results on a national or international scale.

In order to be comparable across borders and over time, measurements need be traceable to appropriate and stated references. CRMs play a key role in implementing the concept of traceability of measurement results in chemistry, biology and physics among other sciences dealing with materials and/or samples. Laboratories use these CRMs as readily accessible measurement standards to establish traceability of their measurement results to international standards. The property values carried by a CRM can be made traceable to SI units or other internationally agreed units during production. This Guide explains how methods can be developed that will lead to well established property values, which are made traceable to appropriate stated references. It covers a very wide range of materials (matrices), ranging from gas mixtures to biological materials, and a very wide range of properties, ranging from chemical composition to physical and immunoassay properties.

The approaches described in this Guide are not intended to be comprehensive in every respect of the production of an RM and the establishment of its property values, including the associated uncertainties. The approaches given in this Guide can be regarded as mainstream approaches for the production and value assignment of large groups of RMs, but appropriate amendments can be needed in a particular case. The statistical methods described exemplify the outlined approaches, and assume, e.g., normally distributed data. In particular when data are definitely not normally distributed, other statistical methods may be preferred to obtain valid property values and associated uncertainties. This Guide describes in general terms the design of projects to produce a CRM.

2 Normative references

The following referenced documents are indispensable for the application 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 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*

ISO Guide 30, *Terms and definitions used in connection with reference materials*

Guide to the expression of uncertainty in measurement. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, 1993[1)]

1) This edition was corrected and reprinted in 1995.

International vocabulary of basic and general terms in metrology. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, 1993

NOTE The "Guide to the expression of uncertainty in measurement" will hereafter be referred to as "GUM", whereas the "International vocabulary of basic and general terms in metrology" will be referred to as "VIM".

3 Terms, definitions and symbols

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO Guide 30 and VIM, together with the following, apply. The symbols to be used are given in Clause 4.

3.1
reference material
RM
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

NOTE 1 RM is a generic term.

NOTE 2 Properties can be quantitative or qualitative (e.g. identity of substances or species).

NOTE 3 Uses can include the calibration of a measurement system, assessment of a measurement procedure, assigning values to other materials, and quality control.

NOTE 4 An RM can only be used for a single purpose in a given measurement.

3.2
certified reference material
CRM
reference material, characterized by a metrologically valid procedure for one or more specified properties, accompanied by a certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability

NOTE 1 The concept of value includes qualitative attributes such as identity or sequence. Uncertainties for such attributes may be expressed as probabilities.

NOTE 2 Metrologically valid procedures for the production and certification of reference materials are given in, among others, ISO Guide 34 and this Guide.

NOTE 3 ISO Guide 31 gives guidance on the contents of certificates.

3.3
property value
⟨of a reference material⟩ value attributed to a quantity representing a physical, chemical or biological property of a (certified) reference material

3.4
characterization
⟨of a reference material⟩ process of determining the property values of a reference material, as part of the certification process

NOTE 1 The characterization process provides the values for the properties to be quantified.

NOTE 2 In batch certifications, the characterization refers to the property values of the batch.

3.5
between-bottle homogeneity
bottle-to-bottle variation of a property of a reference material

NOTE It is understood that the term "between-bottle homogeneity" applies to other types of packages (e.g. vials) and other physical shapes and test pieces.

3.6
within-bottle homogeneity
variation within one bottle of a property of a reference material

3.7
blending
mixing of two or more matrix materials to obtain a material with specific properties

3.8
matrix material
material as sampled from nature, industrial production or elsewhere

EXAMPLES Soil, drinking water, air.

3.9
spiking
adding a known amount of a compound or element to a matrix material

3.10
short-term stability
stability of a property of a reference material during transport under specified transport conditions

3.11
long-term stability
stability of a property of a reference material under specified storage conditions at the CRM-producer

3.12
life time
⟨of a reference material⟩ time interval during which a reference material may be used

3.13
shelf life
⟨of an RM/CRM⟩ time interval during which the producer of the CRM warrants its stability

NOTE The shelf life is equivalent to the period of validity of the certificate, as described in ISO Guide 31.

4 Symbols

A_i bias term (ANOVA)

a number of groups (ANOVA)

B_i bias term (ANOVA)

b number of subgroups (ANOVA)

ε error term (ANOVA)[2)]

k coverage factor

MS mean square (ANOVA)

n number of observations

n_0 (effective) number of (sub)group members (ANOVA)

2) Throughout this Guide, the term error is used in the strict statistical sense, that is the difference between an observed value and its mathematical expectation.

p	number of laboratories in a collaborative study
s_{bb}	between-bottle (in)homogeneity standard deviation
s_{lor}	standard deviation due to lack of repeatability
s_{lts}	long-term (in)stability standard deviation
s_r	repeatability standard deviation
s_{stab}	standard deviation, due to (in)stability
s_{sts}	short-term (in)stability standard deviation
s_{wb}	within-bottle standard deviation
SS	sum of squares (ANOVA)
u_{bb}	standard uncertainty due to between-bottle (in)homogeneity
u_{char}	standard uncertainty due to characterization
u_{CRM}	standard uncertainty of a property value
u_{lts}	standard uncertainty due to long-term (in)stability
u_{sts}	standard uncertainty due to short-term (in)stability
U_{CRM}	expanded uncertainty of a property value
x_{char}	property value as obtained from characterization
x_{CRM}	property value of a CRM
δx_{bb}	error term denoting between-bottle (in)homogeneity
δx_{lts}	error term denoting long-term (in)stability
δx_{sts}	error term denoting short-term (in)stability
x_{ij}	result of a single measurement in the experiment (ANOVA)
μ	population mean (expectation)

NOTE 1 In some clauses, symbols are used to illustrate typical approaches to solve statistical issues in certification projects. These are explained in the text.

NOTE 2 The symbols MS and SS have been adopted from literature, and do not conform the ISO rules with respect to the use of symbols. For clarity however, it is felt that the convention in the scientific literature should prevail.

5 Design of a certification project

5.1 General

The production of a CRM requires a great deal of planning prior to undertaking any actual activity in the project. A substantial part of the planning deals with the amount of material needed, as well as with the design of the homogeneity, stability and characterization studies. The design also includes the choice of appropriate measurement methods for these studies. The number of samples to be produced is a very important variable in the planning process. The number of samples and the amount of raw material depend on all these factors. In the clauses about homogeneity testing (Clause 7), stability testing (Clause 8), and characterization (Clauses 9 and 10), guidance will be provided on how to plan and implement these processes as part of the certification project. A feasibility study may also be part of the project plan.

5.2 Project definition

The planning of a project starts with the definition of what CRM is to be produced. A typical example of such a definition reads as follows:

> "preparation of a soil CRM containing a series of trace elements at relevant content levels for environmental analytical chemistry with an uncertainty associated with the certified values of less than or equal to x %"

This definition covers the project quite well. What is relevant for environmental chemistry may differ from case to case, but it sufficiently narrows the range of materials. Likewise, "soil" also narrows the number of options for the matrix. In all cases it is important to specify what is to be produced. During the design stage of the project, the definition can be specified in more detail. Finally, the target uncertainty specified ensures that the material will be fit for its intended use. For example, uncertainties associated with values of calibration standards should be considerably smaller than uncertainties associated with values of materials for validation of trace environmental analytical methods.

The proper choice of the "stated references" whereto traceability of the property values is established is a major design issue; it strongly depends on what references are available, what is necessary in order for the particular CRM to serve the laboratories performing these measurements routinely, and what is technically feasible. As CRMs are primarily used to make later measurements traceable, the choice of proper references is crucial to the value of the CRM produced, both metrologically and commercially.

The scope for which the CRM is to be used should be stated as well. In most cases, the scope of use is implied by the project definition, but sometimes it needs further elaboration. Such a scope of use does not necessarily exclude other uses, but it should be kept in mind that such uses are not (necessarily) covered by the certificate or documentation provided. The scope for which the RM is to be used can be based on legislation and/or international treaties.

5.3 Transport issues

Prior to starting the actual work, it is important to consider whether the CRM, once ready, can be shipped in agreement with existing regulations. Many CRMs impose a risk with regard to health or safety when people are exposed to the material directly. Proper packaging and appropriate labelling are primary requirements to meet regulations for the transport of (potentially) hazardous goods. Sometimes, legislation or regulations prohibit the transport of materials having certain properties (e.g. viruses, diseases), which may mean that a CRM cannot be sold at all. It is recommended to review all aspects of transport and packing prior to starting the actual certification project.

5.4 Collection of starting material

The first task in a certification project is to obtain a sufficient amount of starting material(s) with the desired properties. For matrix materials, it should be noted that there may be restrictions with respect to the properties of materials. Some material/property combinations are rare, or may be rare in combination with other properties. Often a compromise must be found. In some cases, blending and/or spiking techniques may solve this problem.

The amount of material needed is dictated by the following:

— the number of samples of the (C)RM needed;

— the need for a feasibility study;

— the number of samples needed for the homogeneity study;

— the number of samples needed for the stability study;

— the number of samples needed for the characterization of the candidate CRM;

— the amount of material needed for one measurement.

The required number of samples needed of a candidate CRM is a commercial issue and should be carefully planned beforehand. An important variable is the number of samples likely to be distributed during the lifetime of the CRM. As lifetime is a function of intrinsic stability, this variable also affects the amount of raw material that is needed. For instance, many microbiological materials have limited intrinsic stability and, therefore, their lifetimes are expected to be shorter than, for example, that of a dry sediment certified for trace elements. For an equal number of samples to be dispatched per year, the number of samples needed for the microbiological material is smaller than for the dry sediment. On the other hand for microbiological CRMs, many more samples might be needed for stability testing in the first year(s), or through the entire lifetime of the material.

5.5 Feasibility study

When there are concerns about the feasibility of producing and characterizing a sufficiently homogeneous and stable CRM, a feasibility study may be considered (see Reference [11]). Questions with respect to, for example, the best way of preparing the sample, the stability of the material, or the fitness for purpose, may justify the inclusion of a feasibility study in the project (see References [11], [12]). Sometimes a feasibility study is organized to enable laboratories likely to be involved in the characterization to fine-tune their equipment and their procedures. For a feasibility study aiming at the characterization, it is recommended to have a batch of material slightly different from material used for the candidate CRM.

5.6 Required lifetime and shelf life

The expected lifetime of a reference material is an important variable in the planning of the certification project. Another relevant parameter with respect to the stability is the shelf life of the CRM. Depending on the nature of the mechanisms affecting the stability of the material, various actions may be taken to improve the shelf life and/or lifetime. Adjusting the water activity is one of the first options to be considered, as excessive drying or too high a water content can destabilize the material. In many cases, moisture plays a key role in mechanisms leading to instability of the matrix and/or parameters. In other cases, sterilization or pasteurization of the material might be considered in order to stop bacterial activity. However, these measures can also have a negative effect on stability. Relevant information regarding stability and storage conditions can be found in the literature or can be obtained from users of similar types of materials (industry, etc.). When preparing solutions, additives may increase the shelf life and/or lifetime. The shelf life of a material is a function of the storage conditions as well as a function of the quality of the stability study. The latter determines to what extent the results can be extrapolated (see 8.5).

5.7 Sample preparation

5.7.1 Preamble

It is difficult to give general guidance on the preparation of reference materials. This subclause is intended to give guidance on some specific aspects without the aim of being exhaustive. It is merely a collection of aspects needing careful consideration, which are frequently highly relevant for the success of a certification project.

5.7.2 Synthetic materials

Synthetic reference materials, such as pure substances, solutions and gas mixtures, are prepared in a completely different way from most matrix reference materials. For the preparation of pure substances, purification techniques may be necessary to reduce the total amount of impurities. The choice of these techniques depends on the main component of interest, and may include distillation and/or recrystallization techniques. After a subdivision process (when preparing a batch CRM), the batch should be treated as described in 5.7 to 5.9.

Many solutions and gas mixtures are prepared by means of gravimetry, for which a well-established uncertainty budget can often be obtained. The purity (or composition) of the starting materials enters into the model for calculating the composition of the candidate CRM, as does its uncertainty. For the preparation of batches of materials, volumetric techniques are widely used as well. Usually volumetric methods are somewhat easier to handle, but they are usually also associated with a larger uncertainty than would be the case when prepared gravimetrically.

5.7.3 Blending of materials

Blending of two or more matrices may be considered if a particular property value is considered to be too high or too low. The process is best carried out with matrices of similar kinds, although what is considered to be "similar kinds" may differ widely. For proper blending, the material should be in such a state that agglomeration of particles is suppressed. Usually the moisture content of the materials involved is the dominant factor. If the material is "air-dry", usually (but certainly not always) agglomerates disappear during a good mixing process. The same is true for materials that behave like slurries. There is a potential problem when the agglomerates do not disappear during mixing. Some level of agglomeration of particles may be inevitable. For instance, soy powder with less than 2 % water is still sticky.

A further requirement for proper blending of different matrices is that the densities and the particle size distributions of the materials being blended be sufficiently similar and, for the distribution, sufficiently narrow. This will substantially reduce the segregation risks. With appropriate technology and correct implementation of particle size reduction and blending techniques, it is usually possible to obtain a batch of material that has good properties with respect to homogeneity and stability.

In case of doubt, the blended material may be subjected to a quick homogeneity test, where several portions of the blended material are investigated for homogeneity of the properties to be certified. Such a study may be run on a small number of portions, but a large enough number to obtain some idea about the homogeneity. Typically, 10 portions should be considered to provide meaningful results for taking a decision as to whether the blend material is suited for further processing.

5.7.4 Spiking

There are cases where spiking should be considered as a suitable method for the production of a reference material. Such cases include extracts prepared from solid-state materials. Another example is a series of three CRMs of PCBs in pork fat, where the CRM at elevated temperatures is a liquid. Other examples where spiking is a good method for obtaining CRMs of desired properties are liquids, metals and alloys, oils and workplace atmospheres.

A major problem with spiking is the achievement of sufficient homogeneity and stability of the candidate reference material. Using a proper spiking method can lead to a material that fulfils the requirements with respect to homogeneity and stability, even for solid-state materials. A suitable spiking method for solids is, for example, an "incipient wetness" technique, where the component to be spiked is dissolved in a suitable amount of solvent that is just sufficient to completely wet the surface of the solid. The solvent should be chosen in such a way that its rate of evaporation can be controlled. If the rate of evaporation is too high, the spike may come out again from the pores and cluster. In that case the spike will not be sufficiently well bonded to the surface, which has an impact on the stability of the material. Too low rates of evaporation will lead to migration of other constituents present in the matrix, or even their loss.

For some groups of matrix CRMs, however, spiking is clearly an inappropriate method for obtaining a material with desired values for properties to be certified, as it may lead to CRMs that behave completely differently from normal routine samples. As a rule, major differences in the binding of the naturally incurred and spiked analytes can be expected, leading to differences in, for example, the extraction behaviour. The equivalence of the spiked material to naturally (contaminated) material should therefore be checked to make the material representative of real samples.

5.7.5 Homogenization and subdividing

The sampled material usually undergoes several preparation steps before it becomes a reference material. Necessary steps in this process include drying, particle size reduction, sieving, stabilization and subdividing/bottling. At the design stage of the project, it should be established how far the sample preparation will be extended. For instance, it is possible to prepare a sampled material in such a way that it can be measured directly as an extract. In many cases, however, it is preferable that the sample preparation should leave the sampled material in its original state, although heterogeneity should usually be decreased and stability should be increased as a result of the sample preparation process.

The required uncertainty of the property values of the RM and the required lifetime set requirements with respect to the choice of sample preparation techniques. It should be borne in mind that the way in which the candidate reference material is prepared influences the possible use of the material. For example, distributing an extract will make it impossible to check for the accuracy of the extraction step in the customer's laboratory. Therefore, the objectives of preparing a CRM should be kept in mind when deciding how the raw material is prepared to become suitable to be certified in view of the scope of use of the CRM.

5.8 Homogeneity study

A homogeneity study is necessary in batch certification projects to demonstrate that the batch of bottles (units) is sufficiently homogeneous. Aspects of quality assurance are as important as the determination of the remaining batch between-bottle variation, which is an uncertainty component to be included in the uncertainty estimate of the property value of the CRM. Even when a material is expected to be homogeneous, as in the case of solutions, an assessment of the between-bottle inhomogeneity is required. When dealing with solid-state reference materials, including slurries and sludges, a within-bottle homogeneity study should be foreseen to determine the minimum sample intake. In principle, this homogeneity study does not add to the uncertainty of the property value in question. The number of extra samples needed mainly depends on the between-bottle homogeneity study. The minimum number of bottles selected at random is between 10 and 30, but should generally not be smaller than 10.

The optimal number of samples for a homogeneity study can be determined by statistically supported design techniques. Such methods usually take into consideration the inability of detecting any inhomogeneity, for example due to the uncertainty of the measurements. Furthermore, the number of bottles depends on the batch size, so that the number of samples picked from the batch may be considered to be "representative" of the whole batch. This requirement should be balanced with the uncertainty of the measurements, which is (under repeatability conditions) a function of the repeatability standard deviation of the measurement and the number of replicates. The above-mentioned statistical techniques may be of help with balancing the number of bottles and the number of replicates, so that the best approach is chosen.

5.9 Stability study

Stability testing aims to determine the remaining degree of instability of the candidate RM after preparation, or to confirm the stability of the material. Even "stable" materials may show instability for one or more property values. A distinction is made between the stability under specified

— storage conditions (long-term stability), and

— transport conditions.

As in the case of a homogeneity study, quality assurance aspects are as important as determining the uncertainty budget due to instability effects. The long-term stability concerns the remaining instability of property values of the CRM under specified storage conditions. It is therefore important to specify these conditions accordingly and to study the stability of the material under the same conditions. A reference temperature should be chosen such that it is practically certain that the material is stable at that temperature. Many biological and environmental reference materials show some degree of instability, despite the effort put into defining/determining optimal storage conditions. Transport conditions should ideally be chosen so that the instability of the material during transport does not exceed that of the material on the shelves of the producer. Short-term stability is therefore only relevant as an uncertainty component when the stability of a CRM is affected by the specified transport conditions (e.g. from the producer to the user) in excess of the storage conditions.

The short-term stability study is typically carried out at different temperatures, to study the effect of different temperatures on the properties of the material. Temperatures of samples can vary during transport between −50 °C up to +70 °C, depending on the type of packaging and transport modality. Based on the observed effects, the transport conditions may be defined and packaging instructions drawn up to effectively eliminate any unwanted side effects. A short-term stability study takes typically 1 to 2 months, but may be extended when the optimal storage conditions are to be determined simultaneously.

The stability study requires a considerable number of bottles (units). For each point in time, more than one bottle should preferably be available. As most long-term stability studies last between 24 and 36 months, with typically 5 or 6 points in time, at least 10 to 12 bottles are needed per temperature. When the design foresees multiple temperatures, the number of bottles should be multiplied accordingly. For a short-term stability study, usually 3 to 5 points in time are used, over 2 weeks. Following the same reasoning as for a long-term stability study, the number of bottles should be between 6 and 10 for a short-term stability study per temperature. The inhomogeneity of the material will also influence the number of units needed for the test of stability. If the material is inhomogeneous, it is advantageous to make single determinations on several bottles rather than replicate determinations on a few bottles.

The preferred method for conducting a stability study in a batch certification is to work under repeatability conditions. Otherwise, the estimated uncertainty due to instability is unnecessarily enlarged due to the reproducibility effects in the results during stability testing. Working under repeatability conditions is possible using the isochronous design (see Reference [13]). All samples are kept at a reference temperature at which it is assumed that no instability is encountered (not necessarily the envisaged storage temperature). The samples are subjected to the temperature under test in the stability study and kept at this temperature until all samples have been measured. The points in time are defined by the time elapsed between the moment that they are put at the temperature under test and the moment that they are measured.

For the classical layout (see 8.2), a measurement method should be chosen with good reproducibility properties. Because maintaining good reproducibility of a measurement method is considerably harder than maintaining good repeatability during a single experiment, the isochronous design is advantageous over the classical design. Apart from this aspect, the uncertainty in the assessment using the classical design is in any case greater than in the isochronous case, which means that the shelf life that can be derived from an isochronous stability study will (for a given level of uncertainty) be longer than for a stability study using the classical layout. These advantages compensate well for the disadvantage of having no data during the stability study, in particular for methods with (relatively) poor repeatability and reproducibility. When intermediate data are required, those measurements should be taken independently from the isochronous stability study. When certifying a single artefact, there is no choice but the classical layout.

The experimental design for a stability study, including determination of the optimal number of points in time and the number of samples included, may be based on a statistical design, appreciating, for example, the inability of the measurement method to detect any instability. Furthermore, an empirical model is used as in most stability studies, so the number of points in time should be sufficiently large that a proper assessment of the validity of the model can be carried out. For a linear model for example, which has two parameters (intercept and slope), at least 3 or 4 points are needed, but often more to make a more accurate assessment. For models with more parameters, the number of points (in time) in the stability study should be increased accordingly.

5.10 Choice of measurement methods

The measurement method used for the homogeneity study should have very good repeatability and selectivity. For a stability study where samples are measured on different days, the selectivity and the reproducibility of the measurement method are of primary importance. Therefore, methods for homogeneity and stability studies are not necessarily the same. This is not a problem so long as traceability of the results of the homogeneity and stability studies and characterization to a common reference are established. Such a reference may be a material that is suitable for assessing the various calibrations or results from different measurement methods. Ensuring the traceability of all measurements in a certification project is an important requirement (see, for example, ISO Guide 34 and Reference [14]).

For the characterization of the candidate reference material, especially in the case of matrix reference materials, it is often highly desirable to use multiple methods, and often also multiple laboratories. Both the methods and the performance of the laboratories should represent “state-of-the-art”, and they should be able to make their measurements traceable to the references specified in the design of the project.