BS ISO 33405:2024

BSI Standards Publication

# Reference materials — Approaches for characterization and assessment of homogeneity and stability

Documents
Document Sharing Hub

Full document details: BS ISO 33405:2024

# National foreword

This British Standard is the UK implementation of ISO 33405:2024. It supersedes PD ISO GUIDE 35:2017, which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee RMI/1, Reference Materials.

A list of organizations represented on this committee can be obtained on request to its committee manager.

**Contractual and legal considerations**

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

© The British Standards Institution 2024
Published by BSI Standards Limited 2024

ISBN 978 0 539 26827 0

ICS 71.040.30

**Compliance with a British Standard cannot confer immunity from legal obligations.**

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 May 2024.

**Amendments/corrigenda issued since publication**

| Date | Text affected |
|---|---|

Documents
Document Sharing Hub

Full document details: BS ISO 33405:2024

International Standard

# ISO 33405

First edition
2024-05

# Reference materials — Approaches for characterization and assessment of homogeneity and stability

*Matériaux de référence — Approches pour la caractérisation et l'évaluation de l'homogénéité et la stabilité*

Reference number
ISO 33405:2024(en)

© ISO 2024

Documents
Document Sharing Hub

Full document details: BS ISO 33405:2024

**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

Published in Switzerland



Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

# Contents

Page

**Foreword** ... **vi**
**Introduction** ... **vii**
**1 Scope** ... **1**
**2 Normative references** ... **1**
**3 Terms and definitions** ... **1**
**4 Symbols** ... **2**
**5 Conventions** ... **3**
**6 Overview of reference material production** ... **3**
6.1 General ... 3
6.2 Summary of project design ... 3
6.3 Acquisition of starting material ... 5
6.4 Feasibility studies ... 6
6.5 Reference material processing ... 6
6.6 Homogeneity assessment ... 6
6.7 Stability assessment ... 6
6.8 Choice of measurement procedures ... 7
6.9 Metrological traceability ... 7
6.10 Characterization and uncertainty evaluation ... 7
6.11 Commutability assessment ... 7
6.12 Transport issues ... 7
6.13 Value assignment ... 8
6.14 Stability monitoring ... 8
6.15 Reference materials produced in repeated batches ... 8
**7 Assessment of homogeneity** ... **8**
7.1 Preamble ... 8
7.2 Need for an experimental homogeneity study ... 9
7.3 Properties to be studied ... 10
7.4 Statistically valid sampling schemes ... 10
7.4.1 Minimum number of units for a homogeneity study ... 10
7.4.2 Use of statistical power analysis ... 12
7.4.3 Sampling strategies for a homogeneity study ... 12
7.5 Choice and conduct of the measurement procedure for a homogeneity study ... 13
7.5.1 Choice of measurement procedure ... 13
7.5.2 Conduct of measurements for homogeneity studies ... 13
7.6 Homogeneity study designs ... 15
7.6.1 Objective of a homogeneity study ... 15
7.6.2 The basic homogeneity study design – measurement in a single run ... 16
7.6.3 Randomized block design ... 17
7.6.4 Balanced nested design ... 17
7.6.5 Alternative strategies ... 17
7.7 Evaluating a homogeneity study ... 18
7.7.1 Initial inspection for measurement trends and outliers ... 18
7.7.2 Inspection for processing trends ... 19
7.7.3 Evaluation of the between-unit term – basic design ... 19
7.7.4 Evaluation of the between-unit term – randomized block design ... 20
7.7.5 Evaluation of the between-unit term – balanced nested design ... 21
7.7.6 Other homogeneity designs and alternative estimation methods ... 21
7.8 Insufficient repeatability of the measurement procedure ... 22
7.9 Within-unit homogeneity ... 22
7.9.1 Assessing the need for within-unit homogeneity study ... 22
7.9.2 Testing for significant within-unit heterogeneity ... 22
7.9.3 Assessing minimum sample size ... 24

© ISO 2024 – All rights reserved



Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

7.10 Check for sufficient homogeneity 25
7.11 Uncertainty evaluation from homogeneity studies 26

**8 Assessment and monitoring of stability 26**
8.1 Preamble 26
8.2 Assessment of stability 27
8.2.1 Requirement for stability assessment 27
8.2.2 Types of (in)stability 27
8.2.3 General methods for assessment of stability 28
8.2.4 Need for experimental study of stability 28
8.3 Classification of stability studies 29
8.3.1 General 29
8.3.2 Classification according to conditions of measurement 29
8.3.3 Classification according to stability study duration and conditions 30
8.3.4 Classification by study objective 30
8.3.5 Designs for different storage and treatment conditions 31
8.4 General requirements for effective stability studies 32
8.4.1 Overview of requirements 32
8.4.2 Selection of RM units 32
8.4.3 Suitable measurement procedure(s) for stability studies 32
8.4.4 Appropriate experimental design 32
8.5 Evaluation of stability study results 33
8.5.1 General considerations for stability study data treatment 33
8.5.2 The basic stability study: multiple points in time at a single storage condition 34
8.5.3 Isochronous designs 36
8.5.4 Accelerated stability studies with multiple exposure conditions 38
8.5.5 Additional sources of random variation in stability studies 41
8.6 Action on finding a significant trend in a stability study 41
8.7 Uncertainty evaluation from stability studies 42
8.7.1 General considerations for uncertainty evaluation from stability studies 42
8.7.2 Sources of uncertainty in predicted change over time 42
8.7.3 Estimation of stability uncertainties in the absence of significant trends 43
8.7.4 Evaluation of stability uncertainties in the case of a known significant trend 43
8.8 Estimation of storage lifetime (“shelf life”) from a stability study 44
8.9 Instructions for use related to management of stability 44
8.10 Stability monitoring 44
8.10.1 Requirements for monitoring 44
8.10.2 Choice of initial monitoring point and monitoring interval(s) 45
8.10.3 Experimental approaches and evaluation for stability monitoring 46

**9 Characterization of the material 47**
9.1 Preamble 47
9.2 Establishing metrological traceability 48
9.2.1 Principle 48
9.2.2 Metrological references 48
9.2.3 Types of measurands 49
9.2.4 Effect of sample preparation or pre-treatment 49
9.2.5 Verification of traceability 50
9.3 Characterization using a single reference measurement procedure (as defined in ISO/IEC Guide 99) in a single laboratory 50
9.3.1 Characterization by a reference measurement procedure without direct comparison with a CRM of the same kind 50
9.3.2 Characterization by value transfer from a reference material to a closely matched candidate reference material using a single measurement procedure performed by one laboratory 51
9.3.3 Selection of RM units for single-laboratory characterization 52
9.3.4 Formulation methods 52
9.4 Characterization of a non-operationally defined measurand using two or more methods of demonstrable accuracy in one or more competent laboratories 53
9.4.1 Concept 53

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

9.4.2 Study design 54
9.4.3 Evaluation 55
9.4.4 Single-laboratory multi-method studies 56
9.5 Characterization of an operationally defined measurand using a network of competent laboratories 56
9.5.1 Concept 56
9.5.2 Study setup 57
9.5.3 Evaluation 57
9.6 Purity 57
9.6.1 General 57
9.6.2 Direct determination of purity 57
9.6.3 Indirect determination of purity 58
9.7 Ordinal scales 59
9.8 Nominal properties including identity 59
9.9 Characterization of non-certified values 59

**10 Evaluating measurement uncertainty 59**
10.1 Basis for evaluating the uncertainty of a property value of a CRM 59
10.2 Basic model for a batch characterization 60
10.3 Uncertainty sources 61
10.4 Coverage intervals and factors 61

**Annex A** (informative) **Design and evaluation of studies for the characterization of a method-independent measurand using two or more methods of demonstrable accuracy in one or more competent laboratories 62**

**Annex B** (informative) **Statistical approaches 73**

**Annex C** (informative) **Examples 84**

**Annex D** (informative) **Measurement uncertainty evaluation 94**

**Bibliography 95**

© ISO 2024 – All rights reserved



Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

# 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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity, or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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 334, *Reference materials*.

This first edition cancels and replaces ISO Guide 35:2017, which has been technically revised.

The main changes are as follows:

— technical requirements for the characterization and the assessment of homogeneity and stability of reference materials as stipulated in ISO 17034 is reiterated in this document with additional guidance on approaches that can be used.

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.

© ISO 2024 – All rights reserved


Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

# Introduction

The production of reference materials (RMs) is a key activity for the improvement and maintenance of a worldwide coherent measurement system. As detailed in ISO 33403, RMs with different characteristics are used in measurements, such as calibration, quality control, proficiency testing and method validation, as well as for the assignment of values to other materials. Certified reference materials (CRMs) are also used to confirm or establish metrological traceability to conventional scales, such as the octane number, hardness scales and pH.

To be comparable across borders and over time, measurements need to 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 substances and materials. 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 the International System of Units (SI) or other internationally agreed references during production. This document explains how approaches can be developed that will lead to well established property values, which are made traceable to appropriate stated references.

For RM producers, this document refers to ISO 17034, ISO 33401 and ISO Guide 30 that support the production and certification of RMs:

— ISO 17034 outlines the general requirements to be met by an RM producer to demonstrate competence;

— ISO 33401 describes the contents of certificates for CRMs and of accompanying documents for other RMs;

— ISO Guide 30 contains terms and definitions related to reference materials.

Alongside developments in RM production approaches, the range of classes of RMs is growing with advances in technology, increasing the need for more widely applicable technical guidance in RM production. In addition, increasing use of ISO/IEC 17025 and ISO 15189 by laboratories has led to greater demand for clear statements of metrological traceability.

This document describes examples of possible designs for homogeneity, stability and characterization studies that are in line with ISO 17034. It also contains specific provisions concerning the establishment of metrological traceability in RM production.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

Documents
Document Sharing Hub

# Reference materials — Approaches for characterization and assessment of homogeneity and stability

## 1 Scope

This document explains concepts and provides approaches to the following aspects of the production of reference materials (RMs):

— the assessment of homogeneity;

— the assessment of stability and the management of the risks associated with possible stability issues related to the properties of interest;

— the characterization and value assignment of properties of an RM;

— the evaluation of uncertainty for certified values;

— the establishment of the metrological traceability of certified values.

The guidance given supports the implementation of ISO 17034. Other approaches can also be used as long as the requirements of ISO 17034 are fulfilled.

Brief guidance on the need for commutability assessment (6.11) is given in this document, but no technical details are provided. A brief introduction for the characterization of qualitative properties (9.6 to 9.8) is provided, together with brief guidance on sampling such materials for homogeneity tests (7.4.1.2). However, statistical methods for the assessment of the homogeneity and stability of RMs for qualitative properties are not covered. This document is also not applicable to multivariate quantities, such as spectral data.

NOTE ISO 33406 gives more information on the production of RMs with one or more qualitative property values.

## 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 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 3534-3, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

ISO Guide 30, *Reference materials — Selected terms and definitions*

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

ISO 17034, *General requirements for the competence of reference material producers*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-2, ISO 3534-3, ISO Guide 30, ISO/IEC Guide 99 and the following apply.

© ISO 2024 – All rights reserved



Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

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

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

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

**3.1**
**classical stability study**
experimental study of reference material stability in which individual samples are prepared at the same time and are measured as time elapses

**3.2**
**isochronous stability study**
experimental study of reference material stability in which units exposed to different storage conditions and times are measured in a short period of time

Note 1 to entry: Figure 6 is an illustration of an isochronous stability study.

## 4 Symbols

| | |
|---|---|
| $a$ | number of RM units in a homogeneity study |
| $d$ | measurement bias |
| $k$ | coverage factor or (as subscript) index |
| $L_d$ | a limit of detection (minimum detectable value of the net state variable) calculated using the methods of ISO 11843-1 |
| $N_{min}$ | minimum number of RM units for a homogeneity study for batch sizes over 100 units |
| $N_{prod}$ | number of RM units produced in a single batch |
| $n_r$ | number of runs in a blocked or nested homogeneity study design |
| $p$ | number of laboratory means in an interlaboratory certification exercise |
| $s_{bu}$ | between-unit component of variance from a homogeneity study, expressed as a standard deviation |
| $s_r$ | repeatability standard deviation |
| $s_R$ | within laboratory reproducibility standard deviation |
| $t_{lts}$ | duration of a long-term stability study |
| $U_{CRM}$ | expanded uncertainty associated with a property value of the certified reference material (CRM) |
| $u_{bu}$ | standard uncertainty associated with between-unit variability |
| $u_{char}$ | standard uncertainty associated with a value assigned in a characterization study |
| $u_{CRM}$ | standard uncertainty associated with the property value of the CRM |
| $u_{trg}$ | target measurement uncertainty, expressed as standard uncertainty, for the value of a property to be certified |
| $u_{hom}$ | standard uncertainty associated with heterogeneity |
| $u_{lts}$ | standard uncertainty associated with long-term stability |
| $u_{mon}$ | standard uncertainty associated with a value obtained by measuring an RM at a monitoring point |

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

$u_{trn}$ standard uncertainty associated with the transport stability of the material

$u_{wu}$ standard uncertainty associated with within-unit heterogeneity

$x_{CRM}$ property value of a CRM

$\hat{x}$ estimated value obtained from a robust statistical estimator

$x_{mon}$ value obtained by measuring an RM property value at a monitoring point

$x$ measurement result for a property value

$y_{char}$ value assigned to a RM in a characterization study

$M_{between}$ between-group mean square term from analysis of variance (ANOVA)

$M_{within}$ within-group mean square term from analysis of variance (ANOVA)

# 5 Conventions

In this document, the following conventions are used:

a) A measurand is specified in such a way that there exists a unique 'true value'.

b) All probability assessments described in this document assume normality unless otherwise stated.

c) Throughout this document, the law of propagation of uncertainty is used for the combination of measurement uncertainty contributions. Other methods of evaluating measurement uncertainty can also be applied and, in some cases, it is necessary to do so. Further guidance on these matters is given in ISO/IEC Guide 98-3 and its supplements.

NOTE 1 Variation between units associated with heterogeneity and changes due to instability will possibly not be normally distributed and can result in asymmetric distributions.

NOTE 2 ISO/IEC Guide 98-3 will hereafter be referred to as the “GUM”.

# 6 Overview of reference material production

## 6.1 General

The production and distribution of an RM require careful planning prior to undertaking any actual activity in the project. Subclauses 6.2 to 6.15 provide a brief overview of the steps involved in the production of an RM followed by a description of the main issues involved in planning each step. Detailed guidance on homogeneity assessment, stability assessment and characterization are given in Clauses 7, 8 and 9, respectively.

## 6.2 Summary of project design

The production of an RM involves the following steps:

a) specification of the RM, i.e., the matrix, the properties to be characterized and their desired levels, the intended use of the material and, for CRMs, the target uncertainty[21];

b) design of a procedure for the sourcing of the material;

c) design of an RM manufacturing and/or preparation procedure;

d) packaging and labelling

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

e) selection of measurement procedures appropriate for characterization, homogeneity and stability studies;

f) consideration of metrological traceability for each measured property, particularly for CRMs, for which a statement of metrological traceability is required;

g) assessment of homogeneity;

h) assessment of stability;

i) assessment of commutability (if required);

j) characterization of the RM;

k) combination of the results from homogeneity studies, stability studies and, for CRMs, evaluation of the measurement uncertainties of certified values;

l) preparation of a certificate or product information sheet and, if appropriate, a report on the production and/or certification as per the guidance provided in ISO 33401;

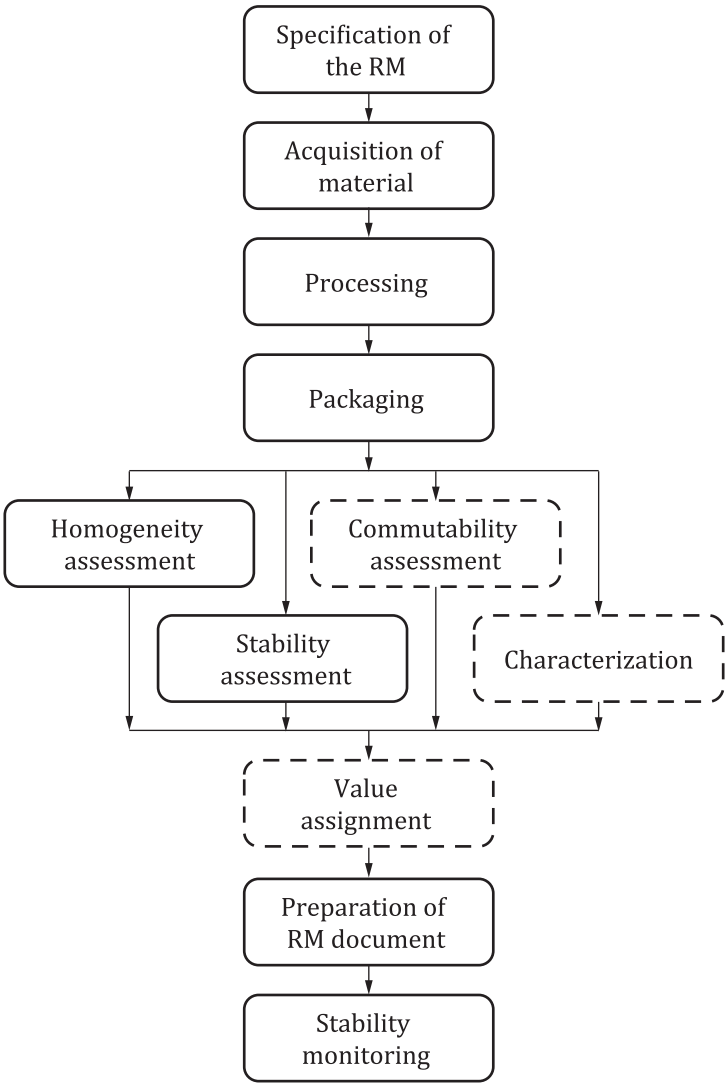
m) specification of storage and transportation conditions and the planned lifespan;

n) post-production monitoring of stability.

The main stages of the project design as listed in 6.2 are shown schematically in Figure 1. Some aspects of the project design are described in more detail in 6.3 to 6.15.

© ISO 2024 – All rights reserved


Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024




NOTE 1 The figure provides a schematic outline of the main steps in producing and maintaining an RM. Boxes with dashed outlines are not always necessary.

NOTE 2 'Packaging' in this figure includes subdivision into individual units in suitable containers for distribution.

**Figure 1 — Schematic outline of a reference material project**

## 6.3 Acquisition of starting material

The first task in an RM production project, after the specification of the RM, is the acquisition of a sufficient amount of starting material(s) with the desired properties. The amount of material needed is determined by the following:

— the number of units of the RM needed for distribution over the expected life of the RM;

— the number of units needed for the homogeneity study;

— the number of units needed for the stability study;

— the number of units needed for the characterization of the candidate RM;

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

— the number of units required for monitoring stability over the expected lifetime of the material;

— the planned size of each RM unit, which has to be sufficient for at least one measurement;

— the need for one or more feasibility studies;

— optionally, additional units to cover contingencies such as, for example, follow-up studies to respond to customer queries, future recertification required by a significant change in the storage conditions or extension of the number of certified properties.

The number of units of an RM that are needed for distribution is often, at least in part, a commercial issue and should be carefully considered before commissioning the collection and processing of the material. In addition, the expected long-term stability of the material in storage can influence the amount of material that can usefully be produced. It is advisable to limit the number of units produced for less stable materials to avoid wastage due to unavoidable degradation over time.

## 6.4 Feasibility studies

Feasibility studies are short studies intended to address concerns about the feasibility of producing and characterizing a sufficiently homogeneous and stable RM. For example, questions such as the best way of preparing the RM or ensuring sufficient stability of the material can be answered by small-scale feasibility studies early in the project[22].

Where characterization is expected to be performed with an interlaboratory study, a feasibility study can identify possible sources of error and enable participants involved in the characterization to optimize their equipment and procedures.

NOTE In a feasibility study intended to test or improve the capabilities of participants in an interlaboratory characterization exercise (see Clause 9), use of a material different from the candidate RM can avoid undue bias in participant results arising from prior knowledge of the candidate RM.

## 6.5 Reference material processing

Processing can involve a range of processes, including, for example:

— synthesis, manufacture or formulation of a synthetic RM;

— drying, lyophilisation, milling, and/or filtration for natural materials;

— addition of stabilizing agents;

— homogenization prior to packaging.

The procedures used depend on the particular material and usually require expert guidance.

## 6.6 Homogeneity assessment

Homogeneity is an important requirement for all RMs and includes both within- and between-unit homogeneity. Between-unit homogeneity is important to ensure that each RM unit carries the same value for each property; within-unit homogeneity is important where subsamples can be taken for measurement by users of the material. Clause 7 gives detailed guidance on homogeneity assessment.

## 6.7 Stability assessment

In accordance with ISO 17034, RMs shall be sufficiently stable for their intended use, so that the end user can rely on the assigned value at any point within the period of validity of the certificate. Typically, it is important to consider stability under long-term storage conditions, under transport conditions and, where applicable, the storage conditions at the RM user's laboratory. This can include consideration of stability after opening if repeated use of the RM is permitted. Clause 8 provides detailed guidance on stability assessment.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

## 6.8 Choice of measurement procedures

In an RM production project, each step that requires measurements can use different measurement procedures because, for example, characterization generally requires minimally biased measurement procedures with low uncertainty; homogeneity studies primarily require the best available repeatability; and classical stability studies typically require measurement procedures that show good precision over time within the same laboratory. The choice of measurement procedures for homogeneity studies, stability studies and characterization are considered in Clauses 7, 8 and 9, respectively.

## 6.9 Metrological traceability

Metrological traceability is key to ensuring the comparability of measurement results over time and between locations, including those used to characterize RMs. By definition, CRMs are accompanied by a statement of metrological traceability for each certified value. The proper choice of the stated references to which metrological traceability of the property values is established, is essential for CRMs, because CRMs are primarily used to make measurement results traceable. Establishment of metrological traceability is considered in detail in 9.2.

## 6.10 Characterization and uncertainty evaluation

Characterization refers to the determination of the property values of the relevant properties of an RM, as part of the production process. Characterization of an RM is described in Clause 9. For CRMs, certified values are accompanied by an associated uncertainty; the evaluation of the uncertainty of the certified value is considered in Clause 10.

## 6.11 Commutability assessment

The commutability of an RM relates to the ability of the RM, characterized by one measurement procedure (usually a reference procedure) to act as a calibrator or quality control (QC) material for a second measurement or testing procedure applied to routine test materials. This is particularly important where different measurement procedures can respond very differently to different types of test materials. Commutability assessment is not required for all RMs but is required for certain classes of RM.

Current ISO/TC 334 information on commutability assessment[23] states that:

“An RM producer should conduct an assessment of commutability where:

a) the intended use requires commutability of calibration or QC materials, and

b) the RM producer warrants that the material is fit for the intended use."

Demonstration of commutability is required when the intended use includes calibration or QC in biological measurement and is not required when the intended use does not include biological measurement and the procedure is known to be adequately specific for the measurand in the matrix of the RM and the intended routine samples.

NOTE 1 It is not necessary to establish commutability when the RM and its origin are obtained from sources and handled the same as samples that would be tested for customers, for example, matrix RMs.

NOTE 2 Further information on commutability assessment can be found in, for example References [24], [25] and [26].

## 6.12 Transport issues

Nearly all RMs are transported to the location of use. The means and conditions of transport of an RM after production are relevant to the need for stability studies (see Clause 8) and it is therefore useful to consider transport conditions at an early stage in the project.

NOTE 1 National and/or international transport regulations can limit the options for transport, prohibit the transport of some materials, or require specific packaging or precautions for safety or other reasons.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

NOTE 2 The time taken for official procedures such as customs or other border control clearance can increase shipment times for some destinations.

### 6.13 Value assignment

Value assignment is the process of combining the results from the homogeneity and stability assessment with the results from the characterization studies to determine the assigned values and their uncertainties. These values are subsequently issued on a certificate or product information sheet.

### 6.14 Stability monitoring

Most RMs are stored for extended periods at the RM producer's premises or by distributors. Since stability assessment cannot usually anticipate all changes that can occur, it is usually necessary, as a part of managing the risks associated with possible instability, to monitor the property values of materials held for extended periods. Because the requirements for monitoring depend in part on knowledge obtained during stability assessment, 8.10 includes guidance on stability monitoring.

### 6.15 Reference materials produced in repeated batches

The need for experimental study of some characteristics (particularly homogeneity, stability, and commutability) can be reduced where the material is produced in a repeat production run following an established procedure. Reliance on prior experience is reasonable so long as:

a) the process for producing batches of the RM has not changed in any way that can adversely affect the end use;

b) the materials used in production of the RM have not changed in any way that can adversely affect the end use;

c) materials previously produced by the same process have shown no failures attributable to the production process, either during routine monitoring or by users;

d) the requirements for the material are reviewed regularly, taking account of the intended use of the material at the time of the review, to ensure that the production process remains fit for purpose.

Consistent performance of the production process should be checked, for example by comparing the property values of samples from successive batches under repeatability conditions.

## 7 Assessment of homogeneity

### 7.1 Preamble

Most RMs are prepared as batches of 'units' (e.g., bottles, vials test items, or test pieces). It is important that that the measurement results for each property value of all distributed units agree with the certified value and its associated uncertainty and, unless sold as single-use units, that the measurement results for each subsample taken from an RM unit also agree with the certified value and its associated uncertainty. ISO 17034 accordingly requires the assessment of the homogeneity of an RM.

Homogeneity can refer either to variation of a property value between separate units of the material or to variation within each unit. It is always necessary to assess the between-unit variation. Where the intended use permits the use of part of a unit – for example, a small portion of a solid or liquid material, or a small region of the surface – it is also usually necessary either to assess the within-unit variability of the material (within-unit heterogeneity) or to provide instructions for use that control the impact of within-unit heterogeneity. These instructions can include, for example, remixing of the sample and, for granular materials, a minimum sample size, because the within-unit heterogeneity is directly reflected in the minimum size of subsample that is representative for the whole unit.

The assessment of homogeneity can include the use of prior evidence (including prior experimental evidence) of the homogeneity of the material, performing an experimental homogeneity study on the candidate RM, or

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

both. In most cases, an experimental study is necessary. Exceptions include, for example, batches of a highly homogeneous material, such as a solution for which previous experimental studies have demonstrated that packaging and storage do not affect the homogeneity; or the production of materials for which each RM unit has a separate assigned value. Subclause 7.2 gives further details about the circumstances and types of material requiring experimental study.

The results of an experimental homogeneity study are usually also used for the calculation of one of the uncertainty components in the certification model (see Clause 10). The magnitude of this uncertainty component can vary widely compared with other components of the uncertainty, depending on the nature of the RM and of the certified property.

To undertake a homogeneity study, a subset of units, typically 10 to 30, is chosen from the batch using a suitable sampling scheme, property values are measured for each unit using a suitable measurement procedure and the results are assessed using appropriate statistical methods to obtain information on, for example, between-unit variability and within-unit variability of the material.

To obtain reliable results, it is important to:

— choose the properties to be studied;

— select a representative subset of units;

— choose a suitable measurement procedure with sufficient repeatability and selectivity;

— make the measurements under suitable conditions following an appropriate experimental design;

— conduct the statistical analysis using valid statistical methods.

Subclauses 7.3 to 7.7 provide guidance on each of these steps. Examples of calculations are provided in Annex C.

Historically, homogeneity studies have tested for statistically significant between-unit differences compared with measurement precision to decide whether a material is homogeneous or not. This approach is not taken in this document; rather, emphasis is placed on deciding whether the between-unit standard deviation is sufficiently small for the intended end use. Statistical tests of significance can, however, be of use in RM production, for example to decide whether further processing is required to reduce heterogeneity to insignificance compared with routine measurement precision. If statistical tests are used, however, the homogeneity experiment should be capable of detecting important heterogeneity, in turn requiring a sufficient combination of precision of the measurement procedure, number of RM units and number of replicates per unit. Statistical power calculations (7.4.2) can assist in ensuring a sufficiently effective test.

## 7.2 Need for an experimental homogeneity study

Materials of natural origin or with complex matrices, such as foodstuffs, soils, ores, and alloys, are typically heterogeneous in composition. Although the magnitude of between-unit differences can sometimes be small or even negligible after homogenization, in other cases, between-unit differences can remain larger than the uncertainty arising from characterization. RMs prepared from such heterogeneous materials should therefore be subjected to an experimental homogeneity study.

RMs prepared as pure compounds or solutions of pure compounds (if certified for purity; not for impurities) are expected to have a high degree of homogeneity. These materials can, however, also show some heterogeneity, for example, due to a density gradient, localized contamination, evaporation of solvent during processing or filling, variations in residual solvent content or metals containing variable amounts of occluded gases. Furthermore, certified values for such materials are often expected to have very small uncertainties, making even a small amount of heterogeneity potentially important. Even in cases where the material is expected to be sufficiently homogeneous for most intended uses, homogeneity should be verified. Verification can include a complete homogeneity study or other check (e.g., a check on melting point consistency between units of a pure organic material).

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

An experimental study of the homogeneity of a material is not essential in the following cases:

— when the material is a repeat production run of a previous material that was produced following the same procedure and that has been shown by experiment to be sufficiently homogeneous; or

— where the production process has been validated and thereby shown to consistently produce sufficiently homogeneous batches of material.

Examples of materials that are sometimes produced in this way include ethanol calibration solutions or elemental calibration solutions prepared by mass and thoroughly mixed to ensure that the mixture will be sufficiently homogeneous for the intended use.

Where assurance of homogeneity relies on a validated production process, QC procedures should be used to confirm consistent operation of the production process. Such procedures can include, for example, operation of a range control chart or standard deviation control chart for monitoring the range or standard deviation of a small number of units measured or criteria for the range of values found in each characterization.

NOTE 1 Subclause 6.15 provides additional guidance regarding reliance on experience gained from previous production batches.

NOTE 2 ISO 7870-2 gives guidance on the use of range control charts and standard deviation control charts.

## 7.3 Properties to be studied

Where homogeneity is to be determined experimentally, it is usually necessary to determine the homogeneity for every property of interest, that is, every property for which the material is claimed to be sufficiently homogeneous for the intended use.

The homogeneity of all properties of interest can be assessed by examining a smaller number of selected properties when:

— there is sufficiently high statistical correlation between particular property values in the type of material of interest to allow useful prediction of one property value from one or more others;

— it can be shown that groups of properties are sufficiently closely associated (e.g., because of their presence in a particular component of a mixture) that measurement of one property in such a group furnishes evidence of homogeneity for other properties in the same group.

It is essential that any subset of properties taken as representing homogeneity for a larger set of properties be appropriately selected based on established chemical or physical relationships. For example, an inter-element concomitance in the mineral phases of an RM would support the assumption that the RM also has a similar degree of homogeneity for the non-selected elements.

In cases where homogeneity is assessed experimentally by using a subset of properties of interest, additional evidence should be gained about the homogeneity of properties that are not studied experimentally. The evidence should be sufficient to show that the uncertainty associated with heterogeneity is not underestimated for those properties that are not experimentally studied.

NOTE Such evidence can be gained, for example, from literature relevant to the material in question, through the stability study or the characterization of the material.

## 7.4 Statistically valid sampling schemes

### 7.4.1 Minimum number of units for a homogeneity study

#### 7.4.1.1 Homogeneity study for quantitative properties

Homogeneity studies for quantitative properties are typically intended to provide information on the variance due to heterogeneity and on any (possibly nonlinear) trends arising from processing. To achieve these objectives, the number of RM units studied should be sufficient to give a reasonable estimate of the

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

between-unit variance and a sufficient number of RM units should be studied to give a clear view of any trends present.

Based on current practice, an acceptable estimate of the between-unit variance for the purposes of uncertainty evaluation can be obtained with nine or more degrees of freedom. For a simple homogeneity study design or a randomized block design (7.6.2 and 7.6.3), this corresponds to the selection of a minimum of 10 units. Where a nested design is to be used for measurements, in which a subset of units is measured in each of several runs, additional units should be included in the study to maintain the required degrees of freedom. Subclause 7.6.4 gives further details about the additional number of units required for a nested design.

Trends arising from processing often appear as an initial trend followed by stable output, as a trend developing late in the process or as a combination of the two. Where these features occur, it is often possible to provide a homogeneous material by discarding the affected units from the beginning or end of the run. However, examining only 10 units will possibly not provide sufficient information about the onset of trends near the ends of a lengthy processing run. Current best practice therefore increases the number of units examined as the total number of units produced, $N_{\text{prod}}$, increases. Typical recommendations are between $\sqrt[3]{N_{\text{prod}}}$ and $3\times\sqrt[3]{N_{\text{prod}}}$ [73]. Taken together with the degrees of freedom requirement, this leads to a recommended minimum number of units $N_{\min}$ for a homogeneity study of materials characterized for a quantitative property given by Formula (1):

$$N_{\min} = \max\left(10, \sqrt[3]{N_{\text{prod}}}\right) \tag{1}$$

where $\max(\ldots,\ldots)$ indicates the maximum of the terms within the parentheses.

NOTE 1 It is not normally useful to examine more than 30 units of an RM characterized for a quantitative property.

NOTE 2 Subclause 7.4.1.3 gives further guidance on the minimum number of units for production batches of 100 or fewer units.

EXAMPLE An RM producer prepares 3 000 units of a candidate RM and intends to undertake a homogeneity study in a single run. Then $N_{\text{prod}}$ = 3 000 and $N_{\min} = \max\left(10, \sqrt[3]{3\,000}\right)$, which is 14,4. This study accordingly requires 15 units of material for the homogeneity study.

#### 7.4.1.2 Homogeneity study for qualitative (“nominal”) properties

For RMs certified for a qualitative, or “nominal” property, the number of units chosen for the homogeneity study should be set based on sampling guidance for inspection by attributes as described in the ISO 2859 series or similar guidance.

Sampling plans for inspection by attributes lead to very high inspection numbers if low proportions of defective units are to be detected by sampling alone. For example, to detect a 1 % defective rate with high confidence based on one or more observed defectives, about 300 randomly chosen units must be inspected (this number is based on a statistical power calculation for 95 % test power for 1 % defectives, assuming batch size much greater than 300 units). This is often unrealistic for typical RM batch sizes. For qualitative materials, therefore, an experimental homogeneity study is likely to be limited to a check for unexpected gross heterogeneity (e.g., greater than 10 % defective), with other information on, for example, origin of material or processing, used to support any statement of homogeneity.

If a certified value of a qualitative property is individually verified for every unit of such a material, it is not necessary to perform a further test of the homogeneity of that property.

#### 7.4.1.3 Small production batches

Some RMs are produced in small batches of 50 or fewer RM units, for example, secondary gas calibration standards. For such small production batches, the minimum number of RM units specified in 7.4.1.1 usually represents a very large fraction of the available RM units. Where the batch size is below a 100 RM units, homogeneity should be assessed on the larger of three RM units or 10 % of the batch size, randomly selected from the batch. Replication should be as high as practically feasible to provide the best available test power

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

for the number of RM units used. Power analysis (7.4.2) can be used to assist in considering desirable replication levels. For example, with three RM units, four observations per unit gives approximately the same power of detecting normally distributed between-unit deviations as observing 10 RM units in duplicate, though with much lower probability of detecting occasional defective RM units in an otherwise homogeneous batch.

### 7.4.2 Use of statistical power analysis

Statistical power analysis can be used to assist in choosing a suitable number of RM units and replicates for the homogeneity study. Power analysis aims to control the probability of failing to detect a particular level of heterogeneity given a proposed statistical test for significant heterogeneity. Power analysis is a specialized topic but is increasingly available in software, some available without charge. The most common example of power analysis in this context is the calculation of the number of RM units and replicate measurements on the assumption that one-way analysis of variance (ANOVA) (see 7.7.3) is to be used to test for a significant between-unit effect.

Although statistical power analysis can be useful in comparing different proposed strategies, considerable care should be taken in its use. In particular, the choice of replicate numbers and unit numbers is very strongly dependent on the assumed distribution of any hypothesized between-sample difference. For example, assuming a normal distribution for the (true) means of different RM units, a common default in power calculation software, leads to a high proposed number of replicates and a small number of RM units. This is a poor choice if the most likely pattern of heterogeneity is a small proportion of discrepant RM units among a largely homogeneous population. In the absence of good information on the likely distribution of different (true) unit means, therefore, power analysis is most useful as an aid in choosing replicate numbers after the proposed number of RM units has been decided.

NOTE 1 When using power analysis, a typical choice of test significance and intended power is 95 % and 80 %, respectively, though note that this provides substantially greater power than the minimum number of RM units and replicates considered at 7.4.1.1.

NOTE 2 Details of some simple power calculations appropriate for RM studies are given in Reference [62].

### 7.4.3 Sampling strategies for a homogeneity study

The sampling scheme used to pick the RM units (bottles, vials test items, or test pieces) for a homogeneity study is typically one of simple random sampling, stratified random sampling or systematic sampling. The general procedures for each, as applied to RM sampling for homogeneity studies, are as follows:

— Simple random sampling takes a random sample from the batch.

— Random stratified sampling typically divides the batch into several segments of equal size (usually by production or packaging sequence or, sometimes, location) and takes an equal number of RM units (often one) at random from each such segment.

— Systematic sampling sets an interval $n_{\text{syst}}$ between sampled RM units and a random start point $n_1$ between 1 and $n_{\text{syst}}$, and takes RM units at $n_1, n_1 + n_{\text{syst}}, n_1 + 2n_{\text{syst}} \ldots$, etc. $n_1$ and $n_{\text{syst}}$ are chosen to give complete coverage of the production batch and the desired number of RM units.

Random samples should be defined using random number generation software or random number tables. Manual arbitrary choice is not equivalent to random selection.

Simple random sampling is most appropriate when there is no known ordering, or for small batches in which a large proportion of the RM units is used for homogeneity study. Stratified random sampling is recommended for most situations, since this guarantees that the RM units picked for the homogeneity study are approximately evenly distributed over the whole batch and avoids coincidence between sampling location and any cyclic effects in processing. Systematic sampling can be applied when there is little risk of overlooking repetitive effects or trends in the batch.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

The sampling scheme should take into consideration potential weaknesses in the method of preparing and storing samples (including, for example, processing and fill trends and possible settling on storage prior to subdivision), thus allowing a critical examination of the prepared batch.

RM units should be numbered before sampling, or numbered in processing order after sampling, to permit subsequent trend analysis.

Alternative sampling schemes can be used where it can be shown that the resulting variance estimates are not biased by the sampling scheme chosen.

## 7.5 Choice and conduct of the measurement procedure for a homogeneity study

### 7.5.1 Choice of measurement procedure

The measurement procedure chosen to make measurements for homogeneity studies should be chosen primarily for good precision during the expected duration of each measurement run (i.e., a good repeatability standard deviation, $s_r$) and, if RM units are to be randomized among several runs, good between-run precision. The standard deviation for measurements should be small compared with the expected uncertainty for the value for each property measured; ideally, the repeatability standard deviation for the homogeneity study procedure should be less than one third of the desired standard uncertainty. That is, if the target measurement uncertainty for the property value is $u_{trg}$ (expressed as standard uncertainty), the repeatability standard deviation for the procedure should ideally conform to Formula (2):

$$\frac{s_r}{\sqrt{n_{al}}} \leq \frac{u_{trg}}{3} \tag{2}$$

where $n_{al}$ is the number of observations on each of $n$ aliquots taken from each subsample of an RM unit or each RM unit.

Where this precision requirement cannot be met in practice, for example where RM units do not allow high replication, every effort should be made to ensure that $\frac{s_r}{\sqrt{n_{al}}}$ does not exceed the target standard uncertainty $u_{trg}$, and consideration should be given to increasing the number of RM units.

NOTE 1 Increasing the number of RM units is not as effective for overcoming the effects of poor precision as is increasing the number of repeated observations (7.5.2) per unit for improving measurement precision, unless the focus is on finding infrequent, highly discrepant RM units. Statistical power calculations (see 7.4.2) can be of use in choosing a suitable number of RM units and replicates to give an experimental design that provides similar assurance to that obtained when the precision requirements in Formula (2) are met.

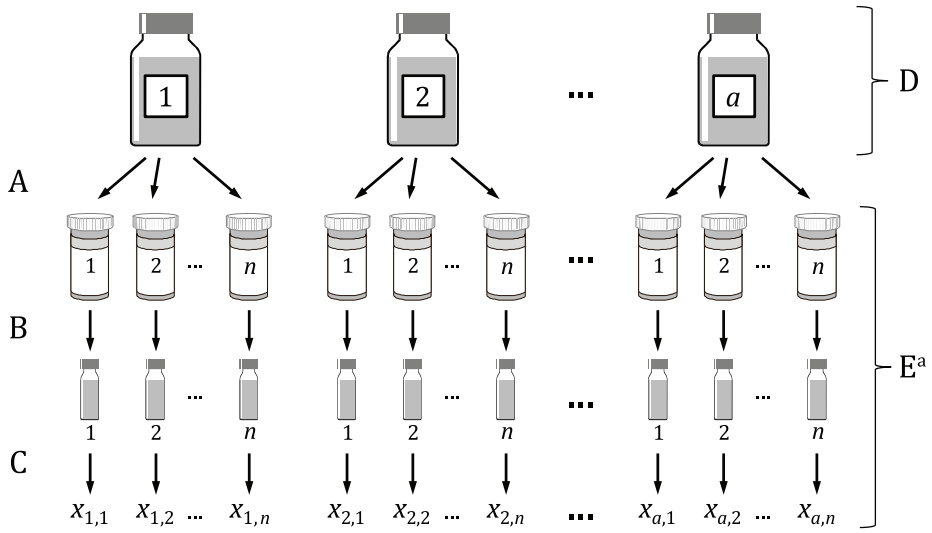
NOTE 2 An additional consideration in the choice of a measurement procedure is the amount and complexity of sample preparation needed to get each specimen into measurable form. Procedures requiring little or no sample preparation (so-called non-destructive methods) can show better precision than procedures requiring extensive, multi-stage, sample preparation.

### 7.5.2 Conduct of measurements for homogeneity studies

It is important, where possible, to compare the dispersion of results for different RM units with the dispersion under repeatability conditions (7.6.1). "Repeatability conditions" involve replication of the complete measurement procedure. It is therefore important that, wherever possible, the complete sample preparation ("transformation") and final measurement step be applied to each subsample taken from each RM unit (see Figure 2). If this is not done, any variation in sample preparation prior to replicate observations will inflate the estimated between-unit term (Figure 3), resulting in an over-estimate of the uncertainty associated with heterogeneity.

© ISO 2024 – All rights reserved


Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024




**Key**

A subsampling

B preparation

C measurement

D contributes to the observed between-unit variation

E operations contributing to observed within-unit variation

[a] $x_{i,j}$ denotes the measurement result for the $j$ th subsample of RM unit $i$.

**Figure 2 — Layout of a between-unit homogeneity study**

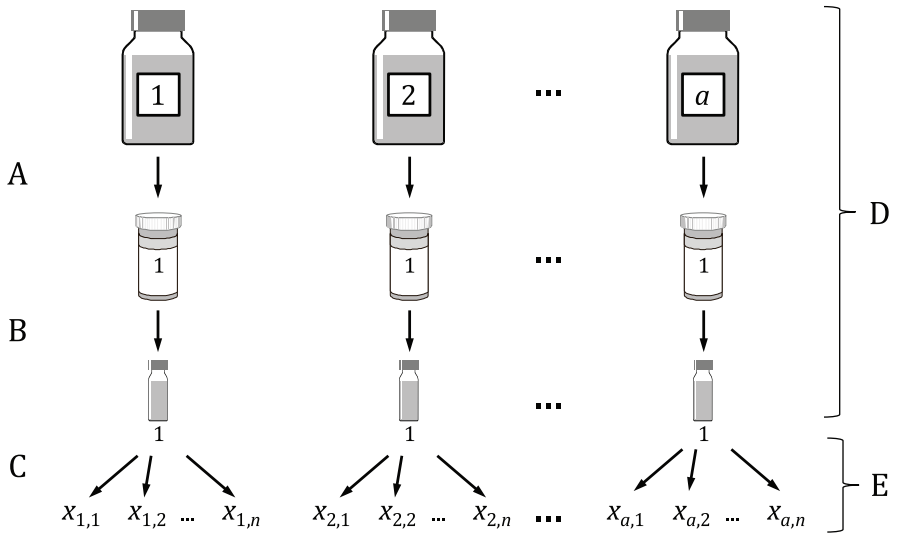
In Figure 2, an ideal case is shown in which subsampling of RM units is possible and multiple ($n$) test portions have been taken from each RM unit sampled from the batch and individually prepared (transformed) for measurement. In this case, the variance within RM units includes the variation due to subsampling, preparation, and measurement. From the perspective of obtaining an unbiased estimate of the between-unit standard deviation, this is the ideal situation because none of these effects contributes to the estimated between-unit variance.

In Figure 3, a layout is shown for the case where subsampling of the RM units is impossible or just not carried out; only one subsample from each of the $a$ RM units is taken and prepared for measurement, and $n$ measurements are made on the prepared test portion. In this design, the effects of sample preparation and any subsampling are included in the observed between-unit variance calculated from the experiment. The calculated within-unit variance includes only the repeatability of the measurement. This procedure will result in a larger estimate of the between-unit variance, in turn over-estimating the between-unit heterogeneity.

It follows that, wherever possible, homogeneity studies should be conducted so that all steps of the measurement procedure are completed on each subsample.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024




**Key**

A subsampling

B preparation

C measurement

D contributes to the observed between-unit variation

E operations contributing to observed within-unit variation

$x_{i,j}$ measurement result for the $j$ th subsample of RM unit $i$.

**Figure 3 — Alternative layout of a between-unit homogeneity study**

## 7.6 Homogeneity study designs

### 7.6.1 Objective of a homogeneity study

The objective of a homogeneity study is to determine the variation of the level of the measurand across the RM batch. A homogeneity study uses statistical evaluation to compare the dispersion of observations on several units of the RM batch with the precision of the measurement procedure to determine the between-unit standard deviation, which can then be used to calculate the uncertainty associated with heterogeneity. This usually involves making an equal number of replicate measurements on each RM unit as illustrated in Figure 2. To obtain the best estimate of the between-unit term, the dispersion of values obtained for RM units in a homogeneity study should be compared with the dispersion of results under repeatability conditions, wherever possible (see Reference [11]). This can be achieved by:

a) conducting all homogeneity measurements for a given property in a single measurement run; or

b) using an experimental design which allows separate estimation of within-run, between-run and between-unit variances; the most useful designs for this purpose are simple randomized block designs and balanced nested designs.

The three designs considered are illustrated in Figure 4. Typical experiments using these different designs are described in more detail in 7.6.2 to 7.6.4. Data analysis is covered in 7.7.3 to 7.7.5, respectively.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

A) Run 1: c a b a f e b g j d h k c h i j k f l d i l g e

B)

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Run 1: | a | e | j | b | i | d | h | c | f | l | g | k |
| Run 2: | g | d | j | f | e | i | k | l | h | b | a | c |

C)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Run 1: | c | a | a | h | b | h | b | c |
| Run 2: | k | i | i | l | e | e | l | k |
| Run 3: | g | j | d | d | f | g | f | j |

**Key**

All illustrations show designs for duplicate observations on 12 RM units designated from “a” to “l”.

A) **Simple randomized design**: a single run with all units observed in duplicate in random order.

B) **Randomized block design**: each run includes one observation on every unit, each run being separately randomized.

C) **Nested design**: four units are randomly allocated to each of three runs and duplicate observations are made on each, again in randomized order. Some observations from the same unit are adjacent by chance; see 7.6.2, NOTE 1.

**Figure 4 — Designs for homogeneity studies**

### 7.6.2 The basic homogeneity study design – measurement in a single run

Where the measurement system is stable for the duration of the observations required, all observations can be acquired in a single measurement run. The within-unit and between-unit standard deviations can then be evaluated using a one-way ANOVA design (see 7.7.3, B.1 and C.1). Measurements should be carried out in such a way that:

— trends in the measurement system are not misinterpreted as differences between RM units;

— a trend in the measurements can, if necessary, be separated from a trend in the batch of samples.

This is best achieved by measuring the replicates of the samples used in the homogeneity study in a fully randomized order. Figure 4 A) shows an illustrative example of such a design.

RM producers can also use a systematic arrangement of replicate observations where it can be shown that:

— reasonably expected patterns of measurement error, including trends and step changes in response within the run, will not cause appreciable bias in estimates of between-unit variance;

— subsequent inspection and data analysis procedures can distinguish processing (including packaging) trends from any measurement trends during the homogeneity study.

A randomly chosen order can, by chance, lead to RM units ordered almost in processing order or with almost all replicates for the same RM unit close together in the measurement sequence. A proposed random ordering which produces obvious strong correlation with processing sequence or close grouping of replicates for most or all RM units should be rejected, and a new random ordering selected.

NOTE 1 A randomized order does not guarantee that individual replicates from the same RM unit are not adjacent. Occasional instances of adjacent replicates are common and acceptable in a fully randomized design.

NOTE 2 A systematic forward-reverse ordering of observations that does not cause appreciable bias in the between-unit term in the presence of a simple linear trend in instrument response with time[28] is illustrated in the following example.

EXAMPLE Suppose 10 samples are used for a homogeneity study, with three replicates. A possible systematic scheme for conducting the measurements reads as follows:

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

| | | |
|---|---|---|
| | Replicate no. 1: | 1 – 3 – 5 – 7 – 9 – 2 – 4 – 6 – 8 – 10 |
| Run: | Replicate no. 2: | 10 – 9 – 8 – 7 – 6 – 5 – 4 – 3 – 2 – 1 |
| | Replicate no. 3: | 2 – 4 – 6 – 8 – 10 – 1 – 3 – 5 – 7 – 9 |

### 7.6.3 Randomized block design

In a randomized block design for $n_0$ replicates on each of $a$ RM units of an RM, the simplest randomized block design involves $n_0$ measurement runs, and each unit is measured once in each run, in random order. Runs should be randomized individually. Figure 4 B) shows an illustrative example of such a design. Two-way ANOVA without replication, or other methods (see 7.7.4 and 7.7.6), can then estimate the within- and between-unit standard deviations independently of the run effect.

A randomized block design is appropriate where the requisite number of replicates on all RM units cannot be included in a single run due to time constraints or instrumental constraints. It is also useful where a measurement process is prone to drift towards the end of longer runs; in that circumstance, better precision (measured as the within-unit standard deviation) can be obtained using several short runs in a randomized block design than can be achieved in a single run.

NOTE 1 In a classical statistical analysis (using ANOVA calculated from sums of squares), the degrees of freedom for the between-unit standard deviation remains equal to $a-1$ for a simple randomized block design, where $a$ is the number of RM units. However, the residual degrees of freedom, which is associated with the within-unit variance, is reduced by one per additional run.

NOTE 2 In a simple randomized block design as described, the calculated within-unit standard deviation includes both variation due to the measurement procedure and variation due to any within-unit heterogeneity.

### 7.6.4 Balanced nested design

In a balanced nested design for $a$ RM units, a number of runs $n_r$ are conducted, and each run includes the same number $a_w$ of RM units, which are each measured with $n_0$ replicates, all replication being carried out within the same run. Each run is randomized separately. Figure 4 C) shows an illustrative example.

Nested designs reduce the degrees of freedom associated with the between-unit standard deviation. For these designs, the number of RM units studied should be increased to retain sufficient degrees of freedom. This usually requires adding a minimum of one additional unit for each additional measurement run in a nested design. In addition, if the statistical method to be used relies on a balanced design, sufficient further RM units should be added to ensure that every measurement run includes the same number of RM units.

NOTE Balanced designs simplify analysis using classical ANOVA based on sums of squares. This makes it advantageous to maintain the same number of RM units per group if ANOVA is to be performed manually or by software (e.g., spreadsheet applications that provide basic ANOVA functionality). More recent variance estimation methods (see 7.7.6) are computationally more complex but much less affected by unbalanced designs in which different numbers of RM units or measurement replicates appear in each group.

EXAMPLE An RM producer using a measurement procedure capable of obtaining up to 12 observations in a single run prepares 2 000 RM units of a candidate RM. Formula (1) yields $N_{\min}$ = 12,6, or 13 RM units, if the study is completed in a single run. Because the RM producer cannot obtain duplicate observations on 13 RM units (i.e., 26 observations) in a single run, the RM producer elects to use a nested design with three runs. This involves adding two additional runs, and the number of RM units is accordingly increased from 13 to 15. The RM producer therefore measures 15 RM units in three groups of five RM units.

### 7.6.5 Alternative strategies

It is not always possible to carry out one of the designs listed in this clause (7.6.2 to 7.6.4). For example, RM units could possibly not permit subdivision for replicate analysis, perhaps because the amount of material provided is sufficient only for a single measurement, because RM units are insufficiently stable once opened, or because the RM units are indivisible (e.g., individual Charpy test pieces). Other situations include very

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

short measurement runs (e.g., three observations per run) or a need to allow for a further source of variation in addition to run effects.

Where RM units cannot be subdivided, the most common strategy is to measure each unit once and to compare the standard deviation (or, more commonly, variance) of those measurements against an independent estimate of the measurement repeatability. For example, the variance of single measurements on 10 RM units can be compared with the variance of 10 observations on a pooled and carefully homogenized preparation of the same material or with a good estimate of precision based on method validation studies. The difference in variance is then taken as an estimate of the between-unit variance.

Where the available run length is too short to permit a complete randomized block or nested design, strategies can include the following:

— Random allocation of replicates among short runs, followed by statistical analysis as for the basic design. This can increase the calculated within-unit variance because the run effect is then included in the within-group term.

— Use of 'balanced incomplete block' designs in which a carefully chosen set of replicates is obtained in each of many short runs. Balanced incomplete block designs are beyond the scope of this document but are described in reference texts on experimental design[27]. Statistical analysis can use classical or other methods.

Where multiple sources of variation are to be controlled, other available designs can be used, including, for example, replicated Latin square designs[28] or other structured experimental designs[27].

## 7.7 Evaluating a homogeneity study

### 7.7.1 Initial inspection for measurement trends and outliers

#### 7.7.1.1 Measurement trends

The experimental data (including any QC observations) should first be inspected for trends in the measurement process. This is most effectively done by visual inspection of a plot of the data in experimental run order. Regression analysis, such as linear regression using a simple straight-line model, can be used to test for a statistically significant trend if an approximately linear trend is apparent.

A measurement trend in a properly randomized homogeneity experiment will generally inflate the within-unit variance, potentially obscuring important between-unit effects. If a significant trend in the measurement results is confirmed, action should be taken to minimize its effects on subsequent analysis. Typical actions include:

— Reject the affected measurement run(s) and repeat the measurements after eliminating the source of the trend.

— Apply a correction to the data based on a suitable smoothing function (e.g., a linear regression) before conducting subsequent ANOVA. Prior correction for a trend reduces the residual degrees of freedom and this should be considered in subsequent variance estimation.

— Include the trend in the data analysis. This procedure adjusts properly for any loss in degrees of freedom, though special statistical software and appropriate statistical expertise are required.

NOTE 1 Examples of these options can be found in Annex C, Clause C.4.

NOTE 2 The RM producer is expected to define criteria to determine whether a trend is significant. In the absence of other criteria, statistical significance at the 95 % level of confidence is usually taken as an indication of a significant measurement trend in homogeneity studies.

NOTE 3 In some cases, for example in a nested or randomized block design with three or more runs, omitting a single failed run can, if the precision in remaining runs is sufficiently good, leave sufficient data for effective homogeneity assessment without repeating the failed run.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

#### 7.7.1.2 Outliers in the basic homogeneity study

Many measurement results are generated as part of a homogeneity study and technical issues can result in occasional outlying results due to instrumental or other attributable causes. Outliers shall not be excluded solely on statistical evidence until they have been investigated and technical grounds can be found to justify exclusion of the outlier. The cause of outliers should be investigated, and appropriate corrective action should be applied. Corrective action in such a case can include, for example:

— removal of a single outlying observation or of all results for a unit of material which shows one such observation (see NOTE 2);

— re-measurement of a suspect unit or subsample within a short period of time;

— repeating a complete measurement run or a complete experiment affected by several outliers.

Outlying unit means shall not be removed unless the technical reason for the deviation has been identified and is attributable to an error in measurement.

NOTE 1 Outliers can be identified by visual inspection or by using outlier tests such as Grubbs' tests or Cochran's test. ISO 5725-2 describes Grubbs' tests. ISO 16269-4 describes graphical outlier tests and other tests for multiple outliers.

NOTE 2 Within-unit outliers can be caused by within-unit heterogeneity. If within-unit heterogeneity is likely, it is prudent to investigate further before discarding any outlying replicate result.

NOTE 3 Omitting a single outlying value from a single unit will result in imbalance. Where the statistical method in use relies on a balanced design, the data for that unit can be omitted provided that the non-outlying values for that unit are consistent with the remainder of the data set and that the number of RM units remaining in the study is sufficient.

NOTE 4 Larger homogeneity studies (e.g., studies involving 20 or more RM units for a given property) can naturally incur more than one within-unit outlier due to instrumental effects. After investigation into the cause of the outlying values, it can be appropriate to discard more than one such observation.

### 7.7.2 Inspection for processing trends

Plotting unit means in processing order provides a simple method of checking for processing run effects. Alternative methods include plots of residuals against unit number after fitting a model that includes measurement run effects; this can be particularly useful after a balanced nested design as unit means can be influenced by measurement run differences. It is not appropriate to apply mathematical corrections to remove trends arising from the processing of the RM. If such a trend is detected and the material is still found to be useful, the uncertainty of homogeneity should be estimated in such a way that it covers the observed trend as well as any random between-unit variation around this trend.

### 7.7.3 Evaluation of the between-unit term – basic design

The basic design described in 7.6.2 can be analysed by one-way ANOVA as described in B.1. The calculation assumes there are $a$ RM units and $n_i$ measurements on RM unit $i$, where $i$ = 1, 2, ..., $a$. Usually, simple homogeneity studies are intended to be balanced designs with the same number $n_0$ of replicate measurements for all RM units, so that $n_1 = n_2 \ldots n_a = n_0$. If a one-way ANOVA approach is then used as described in B.1, the between-unit standard deviation $s_{\mathrm{bu}}$ can be computed from the between-group mean square $M_{\mathrm{between}}$, the within-group mean square $M_{\mathrm{within}}$ and the number of replicates per unit $n_0$ using (with symbols as in Annex B) (see Formula (3) and (4)).

$$x_{ij} = \mu + \delta_i + \varepsilon_{ij} \tag{3}$$

© ISO 2024 – All rights reserved


Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

where

$x_{ij}$ is observation $j$ in group $i$;

$\mu$ is the true mean value of the population of possible results (from which the observed results $x_{ij}$ are assumed to arise);

$\delta_i$ is the (true) deviation of the group $i$ mean from $\mu$;

$\varepsilon_{ij}$ is the random error for observation $j$ in group $i$.

$$s_{\mathrm{bu}}^2 = max\left(\frac{M_{\mathrm{between}} - M_{\mathrm{within}}}{n_0}, 0\right) \tag{4}$$

In this simple case, the between-unit variance $s_{\mathrm{bu}}^2$ is identical to the (squared) between-unit homogeneity contribution to the uncertainty, $u_{\mathrm{bu}}^2$ (see 7.11)[31].

NOTE 1 Annex C, Clause C.1 describes an example of a homogeneity study using one-way ANOVA.

NOTE 2 It can often be assumed that the true RM unit means are normally distributed, though this assumption is not necessary for the calculation given at Formula (4). Likewise, it can often be assumed that the within-unit random measurement error is normally distributed[29].

NOTE 3 If the condition stated in Formula (2) in 7.5.1 is not met, see 7.8.

### 7.7.4 Evaluation of the between-unit term – randomized block design

A randomized block design involving $a$ RM units with $n_0$ replicates, where one of the replicates of each RM unit are measured in each run and $n_r$ runs, can be analysed using two-way ANOVA without replication. The statistical model follows from Formula (4) (see Formula (5))

$$x_{ij} = \mu + A_i + B_j + \varepsilon_{ij} \tag{5}$$

where

$\mu$ is the population mean;

$A_i$ is the effect of unit $i$;

$B_j$ is the effect of run $j$ on the result.

This is identical to the model for the basic design (B.1) except for the addition of $B_j$, the effect of run $j$ on the results. The determination of the between-unit variance $s_{\mathrm{bu}}^2$ is not dependent on the distribution of $A$ or $B$, though large run effects can in principle affect the residual term, indirectly, if the precision depends on mean observed level. Usually, however, run effects that are large enough to show such an effect would invalidate the study entirely. In addition, the analysis assumes that run effects and unit effects are independent and additive. This is also a reasonable approximation for the relatively small effects (compared with the within- and between-group standard deviation) expected in homogeneity studies.

ANOVA applied to a randomized block design, following the model at Formula (5) and with one observation per unit per run, leads to a between-run mean square $M_{\mathrm{B}}$ together with, as before, a between-unit mean square $M_{\mathrm{A}}$, and a residual mean square $M_{\mathrm{within}}$.

The ANOVA table also includes associated degrees of freedom for each term. The residual mean square $M_{\mathrm{within}}$ is an unbiased estimate of the repeatability variance $s_{\mathrm{r}}^2$, although inspection of the ANOVA table will show that the number of degrees of freedom is $(a-1)(n_0-1)$ rather than $a(n_0-1)$ as in the basic

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

design. The between-unit standard deviation $s_{\text{bu}}$ is calculated as in Formula (6) and has $(a-1)$ degrees of freedom.

$$s_{\text{bu}}^2 = max\left(\frac{M_{\text{A}} - M_{\text{within}}}{n_0}, 0\right) \tag{6}$$

NOTE An example of treatment of a randomized block design can be found in Annex C, Clause C.3.

### 7.7.5 Evaluation of the between-unit term – balanced nested design

Variance components for balanced nested designs using $a$ RM units equally distributed among $n_r$ runs and with $m_w$ RM units per run can be calculated using ANOVA methods described in ISO 5725-3, which includes nested designs, or by following standard statistical texts such as Reference [30]. The statistical model is given by Formula (7):

$$x_{ijk} = \mu + A_{ij} + B_j + \varepsilon_{ijk} \tag{7}$$

where

$x_{ijk}$ is the $k$ th observation for unit $i$ ($i$ = 1, 2, ..., $m_w$) in run $j$ ($j$ = 1, 2, ..., $n_r$);

$A_{ij}$ is the effect of unit $i$ in run $j$;

$B_j$ is the effect of run $j$;

$\varepsilon_{ijk}$ ($k$ = 1, 2, ..., $n_0$) is the usual residual error term.

ANOVA for a balanced nested design with RM units nested in runs will again generate a between-run mean square $M_{\text{B}}$ together with a between-unit mean square $M_{\text{A}}$, and a residual mean square $M_{\text{within}}$. Assuming RM units nested within runs and with $n_0$ replicates per RM unit, the between-unit term is again estimated using Formula (4), though with $a-n_r$ degrees of freedom instead of $a-1$ as in the basic design.

Two-way ANOVA for the nested design is described in B.2. For analysing a homogeneity study using $a$ RM units equally distributed among $n_r$ runs and with $m_w$ observations per RM unit, $a$, $n_r$ and $m_w$, respectively, replace $p$, all $b_i$ and all $n_{ij}$ in B.2.

### 7.7.6 Other homogeneity designs and alternative estimation methods

Other designs will usually require analysis using specialized statistical software capable of variance component analysis for a wide range of designs. Some such software continues to use variance estimation based on computing mean squares and equating the mean squares to expected mean squares; this is the method used in 7.7.3 to 7.7.5. Other statistical methods have become available, however; of most interest are maximum likelihood methods[30] and Bayesian analysis[74].

Maximum likelihood methods are among the most general and widely used methods of variance component estimation. The recommended maximum likelihood approach for estimating variances in RM work is the variant known as restricted (or, sometimes, 'residual') maximum likelihood estimation (REML). This provides essentially identical results to those of ANOVA for balanced designs with no missing data, but also deals well with complex designs, missing data, and appreciable imbalance. REML can therefore be used in place of the ANOVA-based methods described as well as in a wider range of circumstances.

Bayesian analysis is increasing in popularity as advanced computational methods become available. Bayesian analysis, however, depends on the use of a prior probability distribution for estimated parameters, including variance parameters. It also generally requires careful specification of the statistical model and in general is more computationally intensive. Although Bayesian methods give good results in experienced hands, they can also give unexpected and misleading results when the model is incorrectly specified or inappropriate priors are chosen, especially for the comparatively small data sets common in RM production.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

Bayesian methods for variance component estimation in RM certification should therefore be used only with the assistance of a qualified statistician experienced in Bayesian analysis of small data sets.

NOTE Detailed calculations for maximum likelihood methods and Bayesian analysis are beyond the scope of this document.

## 7.8 Insufficient repeatability of the measurement procedure

It is not always feasible to perform a homogeneity study with a measurement procedure that is sufficiently repeatable (i.e., that meets the requirement of 7.5)[79]. In those cases, an alternative approach can be used that estimates an upper limit for the between-unit term.

Discussions of various approaches to obtain an uncertainty estimate that accounts for insufficient repeatability of the measurement procedure are given in References [31] and [75]. An example of application of the recommendations of Reference [31] is given in C.2.

## 7.9 Within-unit homogeneity

### 7.9.1 Assessing the need for within-unit homogeneity study

Within-unit homogeneity is an issue that arises when the minimum sample size specified in the instructions for use is smaller than the size of a complete RM unit. It is always important to ensure that the specified minimum sample size is sufficient, and a between-unit homogeneity study alone does not provide such assurance. Where the minimum sample size is small and there is a high risk of appreciable within-unit heterogeneity, an experimental within-unit homogeneity study should be carried out. Situations involving a high risk of appreciable within-unit heterogeneity include one or more of the following:

— minimum sample size is smaller than the unit size;

— a material prepared by mixing powders or granular materials;

— a material prepared by mixing a small quantity of one component into a bulk matrix;

— a material with previously known within-unit heterogeneity.

Experimental assessment can consist of a test for significant within-unit heterogeneity or a determination of the minimum sample size.

### 7.9.2 Testing for significant within-unit heterogeneity

A test for significant within-unit heterogeneity uses a design similar to between-unit homogeneity studies, with the exception that the variance of interest is the within-unit, between-subsample term. Figure 5 shows some possible designs for a within-unit homogeneity study. In the simplest [Figure 5 a)], $m$ test portions, in the size range of the envisaged minimum sample size, are drawn from a single RM unit and each is prepared and measured $n_w \geq 1$ times. Where insufficient aliquots can be drawn from a single RM unit, or to obtain more representative information, more RM units can be included, typically with reduced within-unit replication [Figure 5 b)].

A test for within-unit heterogeneity should provide at least five degrees of freedom for the within-unit term. This can be achieved by, for example, taking six or more subsamples from a single unit [$m \geq 6$ in Figure 5 a)]; taking three subsamples from each of three or more RM units [$a \geq 3$, $m = 3$ in Figure 5 b)]; or by taking duplicate subsamples from each of five or more RM units [$a \geq 5$, $m = 2$ in Figure 5 b)]. Where subsamples are taken from multiple RM units, the analysis should include a between-unit term in a manner analogous to including a run term in the balanced nested design at 7.7.5.

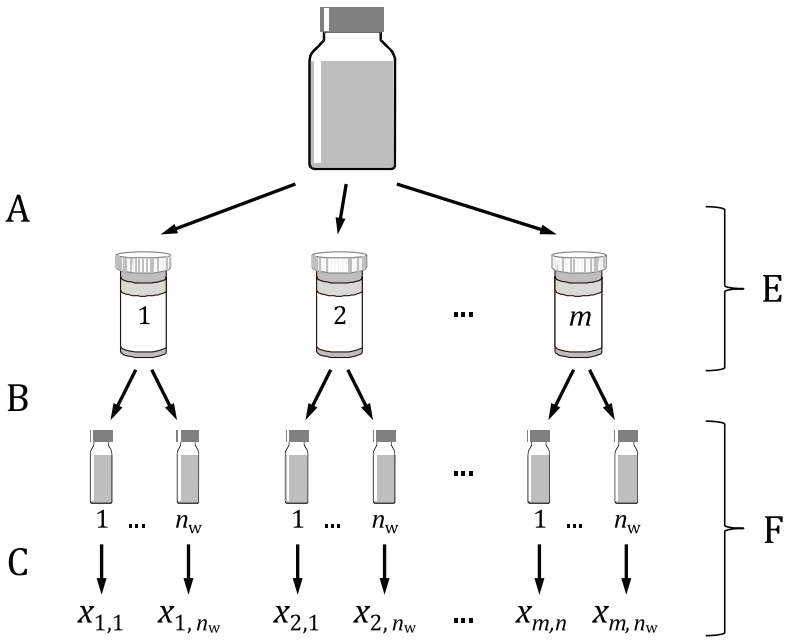
Where replicate analyses can be performed on each subsample, ANOVA followed by an F-test for significant between-subsample variance can be applied. For a single RM unit with aliquots measured multiple times, one-way ANOVA with subsample as the grouping factor can be used as in B.1; where multiple RM units are subsampled and multiple test portions taken, as in Figure 5 b), two-way ANOVA for nested designs is appropriate (B.2). Where only a single observation can be made per subsample ($n_w = 1$), the variance of the

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024

observations (or, from multiple RM units, the within-group term from one-way ANOVA) should be compared with a reliable estimate of the repeatability of the procedure, usually using an F-test for significantly greater variance for results on different subsamples. For cases where the repeatability is well established with high degrees of freedom, such an assessment can be done by comparing the observed variance with the known repeatability variance using a $\chi^2$-test[1].

NOTE Even where it is possible to perform replicate measurements on a subsample, it is not always possible to undertake replicate preparation of a subsample as indicated in Figure 5 b), particularly where the subsample size has been chosen to match the proposed minimum sample size. This results in a design intermediate between those shown in Figure 5 a) and in Figure 5 b).




**a) Within-unit homogeneity study using a single RM unit**

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 33405:2024