

---

**Referenční materiály — návod pro charakterizaci a  
hodnocení homogeneity a stability materiálu**

**TNI  
POKYN ISO 35**

YY YYYY

---

Reference materials -- Guidance for the characterization and the assessment of the homogeneity and stability of the material

Matériaux de référence — Lignes directrices pour la caractérisation et l'évaluation de l'homogénéité et la stabilité de la matière

Tato technická normalizační informace přejímá dokument informativního charakteru, Pokyn ISO 35:2016, vypracovaný v souladu se směrnicemi ISO/IEC, část 1. Překlad byl zajištěn Úřadem pro technickou normalizaci, metrologii a státní zkušebnictví.

## Národní předmluva

### Upozornění na používání této TNI

Pokyn (Guide) je dokument vydaný CEN, CENELEC, ETSI, ISO nebo IEC, obsahující pravidla, zaměření, rady nebo doporučení týkající se evropské a mezinárodní normalizace.

POZNÁMKA Pokyny je dovoleno také zaměřit na problémy, které jsou předmětem zájmu uživatelů evropských a mezinárodních dokumentů.

### Informace pro uživatele

Tato technická normalizační informace stanovuje zásady pro správné zacházení s referenčními materiály

### Citované dokumenty

ISO 3534-2 zavedena v ČSN ISO 3534-2:2010 (010216) Statistika - Slovník a značky - Část 2: Aplikovaná statistika

ISO 3534-3 nezavedena

ISO Guide 30 zaveden jako TNI POKYN ISO 30:2016, Referenční materiály - Vybrané termíny a definice

ISO Guide 34:2009 zaveden jako TNI POKYN ISO 34:2013 (015245) Obecné požadavky na způsobilost výrobců referenčních materiálů

ISO/IEC Guide 98-3 zaveden jako TNI 01 4109-3 (014109) Nejistoty měření - Část 3: Pokyn pro vyjádření nejistoty měření (GUM:1995) (Pokyn ISO/IEC 98-3)

Pokyn ISO/IEC 99 zaveden v TNI 01 0115 Mezinárodní metrologický slovník – Základní a všeobecné pojmy a přidružené termíny (VIM)

### Vysvětlivky k textu této TNI

V případě nedatovaných odkazů na evropské/mezinárodní normy jsou ČSN uvedené v člancích „Informace o citovaných dokumentech“ a „Související ČSN“ nejnovějšími vydáními, platnými v době schválení této normy. Při používání této normy je třeba vždy použít taková vydání ČSN, která přejímají nejnovější vydání nedatovaných evropských/mezinárodních norem (včetně všech změn)

### Vypracování TNI

Zpracovatel: Český institut pro akreditaci, o.p.s., Olšanská 54/3, 130 00 Praha 3, Ing. Eva Klokočnicková, spolupráce: doc. Ing. Zbyněk Plzák, CSc. a doc. RNDr. Václav Sychra, CSc.

Pracovník Úřadu pro technickou normalizaci, metrologii a státní zkušebnictví: Ing. Václav Voves

Technická normalizační komise: TNK XXX Název

**POKYN ISO****Referenční materiály – návod pro charakterizaci a hodnocení homogenity a stability materiálu****ISO GUIDE 35****Český název TNI**

4 vydání

**RRRR-MM-DD**

ICS 71.040.30

**Obsah (bude upraveno při finální redakci)****Contents**

Strana

Page

<b>Obsah (bude upraveno při finální redakci)</b> .....	3
Contents .....	3
Předmluva .....	12
Foreword .....	12
Úvod .....	13
Introduction.....	13
1Předmět technické normalizační informace.....	15
1Scope .....	15
2Citované dokumenty.....	15
2Normative references .....	15
3Termíny a definice.....	16
3Terms and definitions .....	16
4Symboly.....	18
4Symbols.....	18
5Konvence .....	19
5Conventions .....	19
6Přehled výroby referenčního materiálu.....	20
6An overview of reference material production .....	20
6.1 Obecně.....	20
6.1 General.....	20
6.2 Shrnutí návrhu projektu .....	20
6.2 Summary of project design.....	20
6.3 Získání výchozího materiálu .....	22
6.3 Acquisition of starting material.....	22
6.4 Studie proveditelnosti .....	23
6.4 Feasibility studies .....	23
6.5 Příprava referenčního materiálu .....	24
6.5 Reference material preparation .....	24
6.6 Hodnocení homogenity.....	24

6.6 Homogeneity assessment .....	24
6.7 Hodnocení stability .....	24
6.7 Stability assessment .....	24
6.8 Výběr postupů měření .....	24
6.8 Choice of measurement procedures.....	24
6.9 Metrologická návaznost.....	24
6.9 Metrological traceability .....	24
6.10 Charakterizace a vyhodnocení nejistoty .....	25
6.10 Characterization and uncertainty evaluation.....	25
6.11 Hodnocení komutability .....	25
6.11 Commutability assessment.....	25
6.12 Otázky přepravy .....	25
6.12 Transport issues .....	25
6.13 Monitorování stability.....	26
6.13 Stability monitoring .....	26
6.14 Referenční materiály vyráběné v opakovaných šaržích.....	26
6.14 Reference materials produced in repeated batches ..	26
7Hodnocení homogeneity.....	27
7Assessment of homogeneity .....	27
7.1 Úvodní část .....	27
7.1 Preamble .....	27
7.2 Potřeba experimentální studie homogeneity .....	28
7.2 Need for an experimental homogeneity study .....	28
7.3 Vlastnosti, které mají být předmětem studia.....	29
7.3 Properties to be studied.....	29
7.4 Statisticky platná schémata vzorkování .....	30
7.4 Statistically valid sampling schemes.....	30
7.4.1 Minimální počet jednotek pro studii homogeneity .....	30
7.4.1 Minimum number of units for a homogeneity study ...	30
7.4.2 Použití analýzy statistické síly .....	31
7.4.2 Use of statistical power analysis.....	31
7.4.3 Strategie vzorkování u studie homogeneity .....	32
7.4.3 Sampling strategy for homogeneity study.....	32
7.5 Výběr a provádění měřicího postupu u studie homogeneity.....	33
7.5 Choice and conduct of measurement procedure for homogeneity study .....	33
7.5.1 Výběr měřicího postupu.....	33
7.5.1 Choice of measurement procedure .....	33
7.5.2 Provádění měření u studií homogeneity .....	34
7.5.2 Conduct of measurements for homogeneity studies..	34
7.6 Návrhy studie homogeneity .....	37
7.6 Homogeneity study designs .....	37
7.6.1 Cíl studie homogeneity .....	37

7.6.1Objective of a homogeneity study.....	37
7.6.2Základní návrh testování homogenity – měření v jedné sérii .....	38
7.6.2The basic homogeneity test design – measurement in a single run.....	38
7.6.3Znáhodněný blokový návrh .....	39
7.6.3Randomized block design .....	39
7.6.4Vyvážený hierarchický návrh .....	40
7.6.4Balanced nested design .....	40
7.6.5Alternativní strategie.....	40
7.6.5Alternative strategies .....	40
7.7 Vyhodnocení studie homogenity.....	41
7.7 Evaluating a homogeneity study.....	41
7.7.1Prvotní inspekce trendů v měření a odlehlých hodnot .....	41
7.7.1Initial inspection for measurement trends and outliers.....	41
7.7.1.2Odlehlé hodnoty v základní studii homogenity .....	42
7.7.1.2Outliers in the basic homogeneity study .....	42
7.7.2Inspekce trendů ve zpracování.....	43
7.7.2Inspection for processing trends.....	43
7.7.3Vyhodnocení členu mezi jednotkami – základní návrh .....	43
7.7.3 Evaluation of the between-unit term - basic design	43
7.7.4Vyhodnocení členu mezi jednotkami – znáhodněný blokový návrh .....	44
7.7.4Evaluation of the between-unit term - randomized block design .....	44
7.7.5Vyhodnocení členu mezi jednotkami – vyvážený hierarchický návrh .....	44
7.7.5Evaluation of the between-unit term - balanced nested design.....	44
7.7.6Jiné návrhy pro homogenitu a alternativní metody odhadu .....	45
7.7.6Other homogeneity designs and alternative estimation methods .....	45
7.8 Nedostatečná opakovatelnost měřicí metody .....	46
7.8 Insufficient repeatability of the measurement method .	46
7.9 Homogenita uvnitř jednotek.....	46
7.9 Within-unit homogeneity .....	46
7.9.1Posouzení potřeby studie homogenity uvnitř jednotek.....	46
7.9.1Assessing the need for within-unit homogeneity study.....	46
7.9.2Testování významné heterogenity uvnitř jednotky .....	46
7.9.2Testing for significant within-unit heterogeneity .....	46
7.9.3Hodnocení minimálního odběru vzorku .....	50
7.9.3Assessing minimum sample intake.....	50
7.10 Kontrola dostatečné homogenity .....	51

7.10 Check for sufficient homogeneity.....	51
7.11 Vyhodnocení nejistoty ze studie homogeneity.....	51
7.11 Uncertainty evaluation from homogeneity studies .....	51
8Hodnocení a monitorování stability .....	52
8Assessment and monitoring of stability .....	52
8.1 Úvodní část .....	52
8.1 Preamble .....	52
8.2 Hodnocení stability .....	54
8.2 Assessment of stability .....	54
8.2.1Požadavek na hodnocení stability .....	54
8.2.1Requirement for stability assessment.....	54
8.2.2Typy (ne)stability .....	54
8.2.2Types of (in)stability.....	54
8.2.3Obecné metody hodnocení stability.....	55
8.2.3General methods for assessment of stability .....	55
8.2.4Potřeba experimentální studie stability .....	56
8.2.4Need for experimental study of stability .....	56
8.3 Klasifikace studií stability.....	56
8.3 Classification of stability studies .....	56
8.3.1Obecně.....	56
8.3.1General.....	56
8.3.2Klasifikace podle podmínek měření.....	57
8.3.2Classification according to conditions of measurement .....	57
8.3.3Klasifikace podle trvání a podmínek studie stability...58	
8.3.3Classification according to stability study duration and conditions .....	58
8.3.4Klasifikace podle cíle studie .....	58
8.3.4 Classification by study objective .....	58
8.3.5Návrhy různých podmínek skladování a zacházení...59	
8.3.5 Designs for different storage and treatment conditions .....	59
Obecné požadavky na efektivní studie stability .....	60
8.4 General requirements for effective stability studies .....	60
8.4.1Přehled požadavků.....	60
8.4.1Overview of requirements.....	60
8.4.2Výběr jednotek.....	61
8.4.2Selection of units .....	61
8.4.3Vhodná měřicí metoda (metody) pro studie stability..61	
8.4.3Suitable measurement method(s) for stability studies.....	61
8.4.4Vhodný návrh experimentů.....	61
8.4.4Appropriate experimental design .....	61
8.5Vyhodnocení výsledků studie stability .....	63
8.5Evaluation of stability study results.....	63
8.5.1Obecné úvahy o zpracování dat studie stability .....	63

8.5.1	General considerations for stability study data treatment .....	63
8.5.2	Základní studie stability: několik bodů v čase při jedné podmínce skladování .....	64
8.5.2	The basic stability study: multiple points in time at a single storage condition.....	64
8.5.3	Návrhy izochronních studií .....	66
8.5.3	Isochronous designs.....	66
8.5.4	Urychlené studie stability s několika podmínkami expozice .....	69
8.5.4	Accelerated stability studies with multiple exposure conditions .....	69
8.5.5	Další zdroje náhodné variability při studiích stability..	73
8.5.5	Additional sources of random variation in stability studies.....	73
8.6	Opatření při zjištění významného trendu ve studii stability .....	73
8.6	Action on finding of a significant trend in a stability study.....	73
8.7	Vyhodnocení nejistoty ze studie stability .....	74
8.7	Uncertainty evaluation from stability studies.....	74
8.7.1	Obecné úvahy pro vyhodnocení nejistoty ze studií stability .....	74
8.7.1	General considerations for uncertainty evaluation from stability studies.....	74
8.7.2	Zdroje nejistoty předpovídané změny v čase .....	75
8.7.2	Sources of uncertainty in predicted change over time .....	75
8.7.3	Odhad nejistot stability za nepřítomnosti významných trendů.....	76
8.7.3	Estimation of stability uncertainties in the absence of significant trends .....	76
8.7.4	Odhad nejistot stability v případech známého významného trendu.....	76
8.7.4	Estimation of stability uncertainties in the case of a known significant trend.....	76
8.8	Odhad životnosti při skladování (skladovatelnosti) ze studie stability.....	77
8.8	Estimation of storage lifetime ("shelf life") from a stability study.....	77
8.9	Návod k použití ve vztahu k managementu stability ....	77
8.9	Instructions for use related to management of stability	77
8.10	Monitorování stability.....	78
8.10	Stability monitoring .....	78
8.10.1	Požadavky na monitorování stability .....	78
8.10.1	Requirements for monitoring .....	78
8.10.2	Výběr počátečního bodu monitorování a intervalu (intervalů) monitorování.....	78
8.10.2	Choice of initial monitoring point and monitoring interval(s) .....	78
8.10.3	Experimentální návrh.....	80

8.10.3Experimental.....	80
9Charakterizace materiálů.....	82
9Characterization of the material.....	82
9.1 Preamble .....	82
9.1 Preamble .....	82
9.2 Zajištění metrologické návaznosti.....	83
9.2 Establishing metrological traceability.....	83
9.2.1Princip .....	83
9.2.1Principle.....	83
9.2.2Metrologické reference .....	83
9.2.2Metrological references .....	83
9.2.3Typy měřených veličin.....	84
9.2.3Types of measurands .....	84
9.2.4Vliv přípravy vzorku nebo předběžné úpravy.....	85
9.2.4Effect of sample preparation or pre-treatment .....	85
9.2.5Ověření návaznosti.....	85
9.2.5Verification of traceability.....	85
9.3 Charakterizace měřené veličiny nezávislé na postupu jednou metodou v jedné laboratoři .....	86
9.3 Characterization of a method-independent measurand by a single method in a single laboratory .....	86
9.3.1Charakterizace referenční metodou bez přímého porovnání s CRM stejného druhu.....	86
9.3.1Characterization by a reference method without direct comparison with a CRM of the same kind.....	86
9.3.2Charakterizace přímým porovnáním s CRM stejného druhu .....	87
9.3.2Characterization by direct comparison with a CRM of the same kind.....	87
9.3.3Metody přípravy preparátů mísením.....	89
9.3.3Formulation methods.....	89
9.4 Charakterizace měřené veličiny nezávislé na postupu za použití dvou či více metod prokazatelné přesnosti v jedné či více laboratořích.....	90
9.4 Characterization of a method-independent measurand using two or more methods of demonstrable accuracy in one or more laboratories .....	90
9.4.1Koncepte .....	90
9.4.1Concept .....	90
9.4.2Návrh studie .....	91
9.4.2Study design.....	91
9.4.2Vyhodnocení .....	92
9.4.2Evaluation.....	92
9.4.4Studie v jedné laboratoři využívající několik metod ...	94
9.4.4Single-laboratory multi-method studies .....	94
9.5 Charakterizace měřené veličiny definované postupem ve síti kompetentních laboratoří .....	95



9.5 Characterization of an operationally defined measurand using a network of competent laboratories.....	95
9.5.1 Koncepce .....	95
9.5.1 Concept .....	95
9.5.2 Uspořádání studie .....	95
9.5.2 Study setup .....	95
9.5.3 Vyhodnocení .....	95
9.5.3 Evaluation.....	95
9.6 Čistota .....	96
9.6 Purity .....	96
9.6.1 Obecně.....	96
9.6.1 General.....	96
9.6.2 Přímé stanovení čistoty .....	96
9.6.2 Direct determination of purity.....	96
9.6.3 Nepřímé stanovení čistoty .....	96
9.6.3 Indirect determination of purity .....	96
9.7 Identita.....	98
9.7 Identity.....	98
9.7.1 Materiály s certifikací podle původu.....	98
9.7.1 Materials certified based on provenance .....	98
9.7.2 Materiály certifikované na identitu prokázanou měřením.....	98
9.7.2 Materials certified for identity based on measurements .....	98
9.8 Přítomnost/nepřítomnost .....	101
9.8 Presence/absence.....	101
9.9 Pořadové veličiny .....	102
9.9 Ordinal quantities.....	102
9.10 Kvalitativní parametry .....	102
9.10 Qualitative parameters .....	102
9.11 Charakterizace necertifikovaných hodnot.....	102
9.11 Characterization of non-certified values .....	102
10 Vyhodnocení nejistoty měření .....	103
10 Evaluating measurement uncertainty .....	103
10.1 Základ vyhodnocení nejistoty hodnoty vlastnosti CRM.....	103
10.1 Basis for evaluating the uncertainty of a property value of a CRM.....	103
10.2 Základní model pro charakterizaci šarže .....	103
10.2 Basic model for a batch characterization.....	103
10.3 Zdroje nejistoty .....	104
10.3 Uncertainty sources.....	104
10.4 Intervaly pokrytí a koeficienty .....	105
10.4 Coverage intervals and factors.....	105
Příloha A (normativní).....	106
Annex A (normative).....	113

Příloha B (normativní).....	120
Annex B (normative).....	129
Příloha C (informativní).....	138
Annex C (normative).....	146
Příloha D (informativní).....	154
Annex D (informative).....	155
Bibliografie.....	156



## **DOKUMENT CHRÁNĚNÝ COPYRIGHTEM**

© ISO 2016

Veškerá práva vyhrazena. Není-li specifikováno jinak, nesmí být žádná část této publikace reprodukována nebo používána v jakékoliv formě nebo jakýmkoliv způsobem, elektronickým nebo mechanickým, včetně pořizování fotokopíí nebo zveřejnění na internetu nebo intranetu, bez předchozího písemného svolení. O písemné svolení lze požádat buď přímo ISO na níže uvedené adrese, nebo členskou organizaci ISO v zemi žadatele.

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](mailto:copyright@iso.org)

Web [www.iso.org](http://www.iso.org)

Published in Switzerland

## Předmluva

ISO (Mezinárodní organizace pro normalizaci) je celosvětová federace národních normalizačních orgánů (členů ISO). Mezinárodní normy obvykle vypracovávají technické komise ISO. Každý člen ISO, který se zajímá o předmět, pro který byla vytvořena technická komise, má právo být v této technické komisi zastoupen. Práce se zúčastňují také vládní i nevládní mezinárodní organizace, s nimiž ISO navázala pracovní styk. ISO úzce spolupracuje s Mezinárodní elektrotechnickou komisí (IEC) ve všech záležitostech normalizace v elektrotechnice.

Postupy použité při tvorbě tohoto dokumentu a postupy určené pro jeho další udržování jsou popsány ve směrnících ISO/IEC, část 1. Zejména se má věnovat pozornost rozdílným schvalovacím kritériím potřebným pro různé druhy dokumentů ISO. Tento dokument byl vypracován v souladu s redakčními pravidly uvedenými ve směrnících ISO/IEC, část 2 (viz [www.iso.org/directives](http://www.iso.org/directives)).

Upozorňuje se na možnost, že některé prvky tohoto dokumentu mohou být předmětem patentových práv. ISO nelze činit odpovědnou za identifikaci jakéhokoli nebo všech patentových práv. Podrobnosti o jakýchkoli patentových právech identifikovaných během přípravy tohoto dokumentu budou uvedeny v úvodu a/nebo v seznamu patentových prohlášení obdrženy ISO (viz [www.iso.org/patents](http://www.iso.org/patents)).

Jakýkoliv obchodní název použitý v tomto dokumentu se uvádí jako informace pro usnadnění práce uživatelů a neznámá schválení.

Vysvětlení významu specifických termínů a výrazů ISO, které se vztahují k posuzování shody, jakož informace o tom, jak ISO dodržuje principy WTO týkající se technických překážek obchodu (TBT), jsou uvedeny na tomto odkazu URL: Foreword - Supplementary information

Za tento dokument je odpovědná komise ISO/REMCO, Komise ISO pro referenční materiály.

Toto čtvrté vydání zrušuje a nahrazuje třetí vydání (Pokyn ISO:2006).

## Foreword

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

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

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

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

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/REMCO, the Reference Materials Committee of ISO.

This fourth edition cancels and replaces the third edition (ISO Guide 35:2006).

## Úvod

Výroba referenčních materiálů (RM) je klíčovou činností pro zlepšování a podporu uceleného celosvětového měřicího systému. Jak o tom podrobně pojednává Pokyn ISO 33,<sup>[1]</sup> referenční materiály s různými charakteristikami se používají se při měřeních pro kalibrace, řízení kvality a validaci metod a rovněž i pro přidělení hodnot dalším materiálům. Certifikované referenční materiály (CRM) nacházejí kromě toho použití i pro potvrzení nebo zajištění metrologické návaznosti u konvenčních stupnic, jako jsou oktanové číslo, stupnice tvrdosti a pH.

Aby byla měření porovnatelná nehlédě na hranice, a nezávisle na čase musí vykazovat návaznost k vhodným a stanoveným referencím. Certifikované referenční materiály hrají klíčovou roli v koncepci návaznosti měření v chemii, biologii a fyzice z věd zabývajících se látkami a materiály. Laboratoře používají tyto CRM jako snadno dostupné měřicí standardy k zajištění návaznosti výsledků svých měření k mezinárodním standardům. Hodnoty vlastností přenášené CRM se mohou činit návaznými k SI jednotkám nebo jiným mezinárodně uznávaným referencím v průběhu jejich výroby. Pokyn ISO 35 podává výklad jak vyvinout metody, které budou poskytovat dobře stanovené hodnoty vlastností, které budou vykazovat návaznost k odpovídajícím stanoveným referencím.

Pro potřeby výrobců referenčních materiálů existují tři pokyny ISO, ze kterých výroba a certifikace RM může vycházet, aby se zajistilo, že kvalita RM bude vyhovovat požadavkům koncových uživatelů.

- Pokyn ISO 34<sup>[2]</sup> nastiňuje hlavní požadavky, které výrobce referenčních materiálů musí splňovat, aby prokázal svou způsobilost.
- Pokyn ISO 35 poskytuje konkrétnější směrnici zabývající se technickými otázkami a podává základní představu o procesech, jako jsou hodnocení homogenity, stability a přidělení hodnoty u certifikace referenčních materiálů.
- Pokyn ISO 31<sup>[3]</sup> popisuje formát a náplň certifikátů certifikovaných referenčních materiálů a doprovodných dokumentů k dalším referenčním materiálům.

Od té doby co v roce 2006 vyšel Pokyn ISO 35, došlo k revizi Pokynu ISO 34 a nyní se tento při řešení technických otázek ve větší míře opírá o Pokyn ISO 35. Kromě toho, vývoj výrobních metod referenčních materiálů zvýšil rozsah technik pro jejich výrobu, zatímco zvýšené používání Pokynu ISO 34 vyvolalo potřebu technické směrnice širšího rozsahu, která by zahrnovala větší rozsah tříd referenčních materiálů. Navíc čtenější používání ISO/IEC 17025 and ISO 15189 v laboratořích vedlo k vyšší potřebě jasných prohlášení o metrologické návaznosti.

## 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 Guide 33,<sup>[1]</sup> 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 SI units or other internationally agreed references during production. ISO Guide 35 explains how methods can be developed that will lead to well established property values, which are made traceable to appropriate stated references.

For reference material producers (RMPs), there are three ISO Guides that support the production and certification of RMs to ensure that the quality of the RMs meet the requirements of the end users.

- ISO Guide 34<sup>[2]</sup> outlines the general requirements to be met by an RMP to demonstrate competence.
- ISO Guide 35 provides more specific guidance on technical issues and explains the concepts for processes such as the assessment of homogeneity, stability, and value assignment for the certification of RMs.
- ISO Guide 31<sup>[3]</sup> describes the format and contents of certificates for CRMs, and of accompanying documents for other RMs, respectively.

Since the release of ISO Guide 35 in 2006, ISO Guide 34 has been revised and now relies more heavily on ISO Guide 35 for technical guidance. In addition, developments in RM production methods have increased the range of techniques used in RM production, while increasing use of ISO Guide 34 has established a need for more widely applicable technical guidance covering a greater range of classes of RM. In addition, increasing use of ISO/IEC 17025 and ISO 15189 by laboratories has led to greater demand for clear statements of metrological traceability.

Toto vydání Pokynu ISO 35 poskytuje podrobný návod pro širší sortiment návrhů jak provádět studie homogeneity a popisuje ve větším rozsahu strategie zvládnutí stability. Obsahuje rovněž specifická opatření týkající se zajištění metrologické návaznosti při výrobě referenčních materiálů a větší množství postupů přidělování hodnot.

This edition of ISO Guide 35 provides detailed guidance on a larger range of homogeneity study designs, and describes a wider range of stability management strategies. It also contains specific provisions concerning the establishment of metrological traceability in RM production and a wider range of value assignment procedures.

## 1 Předmět technické normalizační informace

Tento Pokyn objasňuje zásady a poskytuje přístupy k následujícím stránkám výroby referenčních materiálů:

- posuzování homogenity,
- posuzování a řízení rizika spojeného se stabilitou materiálu,
- přidělení hodnot vlastností referenčnímu materiálu,
- vyhodnocení nejistoty měření certifikovaných hodnot,
- zajištění metrologické návaznosti certifikovaných hodnot vlastností.

Poskytované návody podporují používání Pokynu 34.<sup>[2]</sup> Mohou se používat i jiné přístupy pokud splňují požadavky Pokynu ISO 34.

Tento pokyn zahrnuje i stručné pojednání o potřebě hodnocení komutability, ale bez jakýchkoli technických podrobností. Poskytuje i stručný úvod do charakterizace kvalitativních vlastností. Nezahrnuje však hodnocení homogenity a stability referenčních materiálů pro kvalitativní vlastnosti. Tento pokyn nelze též použít pro mnohorozměrné veličiny, jako jsou spektrální data.

## 2 Citované dokumenty

V tomto dokumentu jsou uvedeny normativní odkazy na následující citované dokumenty nebo jejich části, které jsou nezbytné pro jeho použití. U datovaných odkazů platí pouze citovaná vydání. U nedatovaných odkazů platí poslední vydání dokumentu (včetně všech změn).

ČSN ISO 3534-2 Statistika - Slovník a značky - Část 2: Aplikovaná statistika

ISO 3534-3 Statistika - Slovník a značky - Část 3: Navrhování experimentů

Pokyn ISO 30, Referenční materiály – vybrané termíny a definice

Pokyn ISO 34:2009 Obecné požadavky na způsobilost výrobců referenčních materiálů

Pokyn ISO/IEC 98-3 Nejistoty měření - Část 3: Pokyn pro vyjádření nejistoty měření (GUM:1995)

Pokyn ISO/IEC 99 Mezinárodní metrologický slovník– Základní a všeobecné pojmy a přidružené termíny (VIM)

POZNÁMKA „Pokyn pro vyjádření nejistoty měření“ bude v dalším odkazován jako GUM, zatímco „Mezinárodní metrologický slovník“ bude odkazován jako „VIM“.

## 1 Scope

This Guide explains concepts and provides approaches on the following aspects of the production of reference materials:

- the assessment of homogeneity;
- the assessment and management of the risk associated with the stability of the material;
- the assignment of values to properties of a reference material;
- the evaluation of measurement uncertainty for certified values;
- the establishment of the metrological traceability of the certified property values.

The guidance given supports the implementation of ISO Guide 34.<sup>[2]</sup> Other approaches may also be used as long as the requirements of ISO Guide 34 are fulfilled.

Brief guidance on the need for commutability assessment is given in this Guide, but no technical details are provided. A brief introduction for the characterization of qualitative properties is provided. However, the assessment of the homogeneity and stability of reference materials for qualitative properties is not covered. This Guide is also not applicable to multivariate quantities, such as spectral data.

## 2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. 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 Guide 34:2009, *General requirements for the competence of reference material producers*

ISO/IEC Guide 98-3, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)*

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

NOTE The “Guide to the expression of uncertainty in measurement” will hereafter be referred to as the “GUM”,

whereas the “International vocabulary of metrology” will be referred to as the “VIM”.

### 3 Termíny a definice

Pro účely tohoto dokumentu platí termíny a definice uvedené v Pokynu ISO 30, Pokynu ISO/IEC 99 a v ISO:3534-2 ISO:3534-3 a termíny a definice níže uvedené.

#### 3.1 referenční materiál RM

materiál, dostatečně homogenní a stabilní vzhledem k jedné či více specifikovaným vlastnostem, u kterého se stanovilo, že se hodí pro jeho zamýšlené použití v měřicím procesu

POZNÁMKA 1 k heslu: RM je generický termín.

POZNÁMKA 2 k heslu: Vlastnosti mohou být kvantitativní nebo kvalitativní, např. identita látek nebo specií (druhů).

POZNÁMKA 3 k heslu: Použití může zahrnovat kalibraci měřicího systému, posuzování měřicího postupu, přidělení hodnot jiným materiálům a řízení kvality

POZNÁMKA 4 k heslu: Pokyn ISO/IEC 99:2007 <sup>[4]</sup> obsahuje analogickou definici (5.13), ale omezuje použití termínu „měření“ jen na kvantitativní hodnoty. Avšak poznámka 3 Pokynu ISO/IEC 99:2007, 5.13 (VIM) výslovně zahrnuje kvalitativní vlastnosti, které nazývá „jmenovitými vlastnostmi“.

[ZDROJ: ISO Pokyn 30:2015, 2.1.1]

#### 3.2 certifikovaný referenční materiál CRM

referenční materiál (RM) charakterizovaný metrologicky platným postupem pro jednu či více specifikovaných vlastností, doprovázený certifikátem referenčního materiálu, který poskytuje hodnotu specifikované vlastnosti, její přidruženou nejistotu a prohlášení o metrologické návaznosti

POZNÁMKA 1 k heslu: Pojem hodnoty zahrnuje jmenovitou vlastnost nebo kvalitativní znak jako je identita či sekvence. Nejistoty pro takové znaky mohou být vyjádřeny jako pravděpodobnosti nebo konfidenční úrovně.

POZNÁMKA 2 k heslu: O metrologicky platných postupech pro výrobu a certifikaci referenčních materiálů pojednávají, mezi jinými, Pokyny ISO 34 <sup>[2]</sup> a 35.

POZNÁMKA 3 k heslu: : Pokyn ISO 31 <sup>[3]</sup> je směrnici o obsahu certifikátů RM.

POZNÁMKA 4 k heslu: Pokyn ISO/IEC 99: 2007 <sup>[4]</sup> obsahuje analogickou definici (5.14).

[ZDROJ: Pokyn ISO 30:2015, 2.1.2]

### 3 Terms and definitions

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

#### 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 to entry: RM is a generic term.

Note 2 to entry: Properties can be quantitative or qualitative, e.g. identity of substances or species.

Note 3 to entry: Uses may include the calibration of a measurement system, assessment of a measurement procedure, assigning values to other materials, and quality control.

Note 4 to entry: ISO/IEC Guide 99:2007 <sup>[4]</sup> has an analogous definition (5.13), but restricts the term “measurement” to apply to quantitative values. However, Note 3 of ISO/IEC Guide 99:2007, 5.13 (VIM), specifically includes qualitative properties, called “nominal properties”.

[SOURCE: ISO Guide 30:2015, 2.1.1]

#### 3.2 certified reference material CRM

reference material (RM) characterised by a metrologically valid procedure for one or more specified properties, accompanied by an RM certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability

Note 1 to entry: The concept of value includes a nominal property or a qualitative attribute such as identity or sequence. Uncertainties for such attributes may be expressed as probabilities or levels of confidence.

Note 2 to entry: Metrologically valid procedures for the production and certification of RMs are given in, among others, ISO Guides 34 <sup>[2]</sup> and 35.

Note 3 to entry: ISO Guide 31 <sup>[3]</sup> gives guidance on the contents of RM certificates.

Note 4 to entry: ISO/IEC Guide 99:2007 <sup>[4]</sup> has an analogous definition (5.14).

[SOURCE: ISO Guide 30:2015, 2.1.2]



**3.3****model měření**

matematický vztah mezi všemi veličinami, které zahrnuje měření

[ZDROJ: Pokyn ISO/IEC 99:2007 <sup>[4]</sup>]

**3.4****hodnota vlastnosti**

<referenčního materiálu (RM)> hodnota odpovídající veličině reprezentující fyzikální, chemickou nebo biologickou vlastnost RM

[ZDROJ: Pokyn ISO 30:2015.2.2.1]

**3.5****certifikovaná hodnota**

hodnota, přidělená vlastnosti referenčního materiálu (RM), která je doprovázena vyjádřením o nejistotě a prohlášením o metrologické návaznosti a uvedena jako taková v certifikátu RM

[ZDROJ: Pokyn ISO 30:2015, 2.2.3]

**3.6****indikativní hodnota**

informační hodnota

informativní hodnota

hodnota veličiny nebo vlastnosti referenčního materiálu, která se poskytuje pouze pro informaci

POZNÁMKA 1 k heslu: Indikativní hodnota se nemůže použít jako reference v řetězci metrologické návaznosti.

[ZDROJ: Pokyn ISO 30:2015, 2.2.4]

**3.7****kalibrátor**

referenční materiál použitý pro kalibraci zařízení nebo měřicího postupu.

[ZDROJ: Pokyn ISO 30:2015, 2.1.21]

**3.8****materiál pro řízení kvality**

referenční materiál používaný pro řízení kvality měření

[ZDROJ: Pokyn ISO 30:2015, 2.1.22]

**3.9****izochronní test stability**

testování stability referenčního materiálu, při níž jednotky vystavené různým podmínkám skladování a po různou dobu jsou změřeny během krátké doby

**3.10****výroba**

<referenčního materiálu (RM)> všechny potřebné činnosti a úkoly vedoucí k uvedení a technickému zabezpečení RM (certifikovaných nebo necertifikovaných)

POZNÁMKA 1 k heslu: Tyto činnosti zahrnují např. plánování, kontrolu, manipulaci s materiálem a skladování, zpracování materiálu, posuzování homogenity a stability, charakterizaci, přidělení hodnot vlastností a jejich nejistot, schvalování a vydávání certifikátů RM nebo jiných prohlášení.

**3.3****measurement model**

mathematical relation among all quantities known to be involved in a measurement

[SOURCE: ISO/IEC Guide 99:2007 <sup>[4]</sup>]

**3.4****property value**

<of a reference material (RM)> value corresponding to a quantity representing a physical, chemical or biological property of an RM

[SOURCE: ISO Guide 30:2015, 2.2.1]

**3.5****certified value**

value, assigned to a property of a reference material (RM) that is accompanied by an uncertainty statement and a statement of metrological traceability, identified as such in the RM certificate

[SOURCE: ISO Guide 30:2015, 2.2.3]

**3.6****indicative value**

information value

informative value

value of a quantity or property, of a reference material, which is provided for information only

Note 1 to entry: An indicative value cannot be used as a reference in a metrological traceability chain.

[SOURCE: ISO Guide 30:2015, 2.2.4]

**3.7****calibrant**

reference material used for calibration of equipment or a measurement procedure

[SOURCE: ISO Guide 30:2015, 2.1.21]

**3.8****quality control material**

reference material used for quality control of a measurement

[SOURCE: ISO Guide 30:2015, 2.1.22]

**3.9****isochronous stability test**

test of reference material stability in which units exposed to different storage conditions and times are measured in a short period of time

**3.10****production**

<of a reference material (RM)> all necessary activities and tasks leading to the release and maintenance of an RM (certified or non-certified)

Note 1 to entry: Activities include, e.g. planning, control, material handling and storage, material processing, assessment of homogeneity and stability, characterization, assignment of property values and their uncertainties, authorization and issue of RM certificates or other statements.

[ZDROJ: Pokyn ISO 30:2015, 2.3.7]

#### 4 Symboly

$a$	počet jednotek referenčního materiálu při studii homogeneity
$d$	vychýlení (bias) měření
$k$	koeficient vychýlení nebo dolní index
$L_d$	mez detekce (minimální detekovatelná hodnota redukované stavové proměnné) vypočtená pomocí metod ISO 11843
$N_{\min}$	minimální počet jednotek RM studie homogeneity u velikosti šarží přesahujících 100 jednotek
$N_{\text{prod}}$	počet jednotek RM vyráběných v rámci jedné šarže
$n_r$	počet sérií v blokovém (blocked) nebo hierarchickém (nested) návrhu studie homogeneity
$p$	počet laboratorních průměrů nebo v mezilaboratorním certifikačním experimentu
$s_{bb}$	složka variability mezi jednotkami u studie homogeneity, vyjádřená jako směrodatná odchylka
$s_r$	směrodatná odchylka opakovatelnosti
$s_i$	směrodatná odchylka za mezilehlých podmínek měření
$s_R$	směrodatná odchylka reprodukovatelnosti
$t_{\text{its}}$	trvání dlouhodobé studie stability
$U_{\text{CRM}}$	rozšířená nejistota přidružená hodnotě vlastnosti CRM
$u_{bb}$	standardní nejistota přidružená mezijednotkové variabilitě
$u_{\text{char}}$	standardní nejistota přidružená hodnotě přiřazené při charakterizaci
$u_{\text{CRM}}$	standardní nejistota přidružená hodnotě vlastnosti CRM
$u_d$	požadovaná standardní nejistota hodnoty vlastnosti, která je předmětem certifikace
$u_{\text{hom}}$	standardní nejistota přidružená heterogenitě
$u_{\text{its}}$	standardní nejistota přidružená dlouhodobé stabilitě
$u_{\text{mon}}$	standardní nejistota přidružená hodnotě získané měřením RM na místě monitorování
$u_{\text{tm}}$	standardní nejistota přidružená přepravní stabilitě materiálu

[SOURCE: ISO Guide 30:2015, 2.3.7]

#### 4 Symbols

$a$	number of reference materials 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
$N_{\min}$	minimum number of RM units for a homogeneity study for batch sizes over 100 units
$N_{\text{prod}}$	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 or in an interlaboratory certification exercise
$s_{bb}$	between-unit component of variance from a homogeneity study, expressed as a standard deviation
$s_r$	repeatability standard deviation
$s_i$	standard deviation under intermediate conditions of measurement
$s_R$	reproducibility standard deviation
$t_{\text{its}}$	duration of a long term stability study
$U_{\text{CRM}}$	expanded uncertainty associated with property value of the CRM
$u_{bb}$	standard uncertainty associated with between unit variability
$u_{\text{char}}$	standard uncertainty associated with a value assigned in a characterization study
$u_{\text{CRM}}$	standard uncertainty associated with property value of the CRM
$u_d$	desired standard uncertainty for the value of a property to be certified
$u_{\text{hom}}$	standard uncertainty associated with heterogeneity
$u_{\text{its}}$	standard uncertainty associated with long term stability
$u_{\text{mon}}$	standard uncertainty associated with a value obtained by measuring a RM at a monitoring point
$u_{\text{tm}}$	standard uncertainty associated with the transportation stability of the material
$u_{\text{wb}}$	standard uncertainty associated with within-unit heterogeneity

$u_{wb}$	standardní nejistota heterogenitě uvnitř jednotky	přidružená	$x_{CRM}$	property value of a CRM
$x_{CRM}$	hodnota vlastnosti CRM		$\hat{x}$	estimated value obtained from a robust statistical estimator
$\hat{x}$	odhadovaná hodnota statistického odhadu	robustního	$x_{mon}$	value obtained by measuring a RM property value at a monitoring point
$x_{mon}$	hodnota získaná měřením vlastnosti RM na místě monitorování	hodnoty	$x$	amount-of-substance fraction
$x$	látkový zlomek		$y_{char}$	value assigned to a reference material in a characterization study
$y_{char}$	hodnota přidělená referenčnímu materiálu v charakterizační studii			

Další symboly používané v příslušných člancích textu budou definovány při jejich prvním použití.

Additional symbols used in particular subclauses are defined on first use in the text.

## 5 Konvence

V tomto pokynu se používají následující konvence.

- a) Měřená veličina je specifikována takovým způsobem, že existuje jediná pravá hodnota'.
- b) Všechny statistické metody používané v tomto pokynu vycházejí z následujících předpokladů.
  - 1) Certifikovaná hodnota je nejlepším odhadem pravé hodnoty vlastnosti CRM.
  - 2) Všechna variabilita, ať už spojená s materiálem (tj. homogenita a stabilita) nebo měřícím procesem je náhodná a odpovídá normálnímu rozdělení pravděpodobnosti nebo může být na toto rozdělení převedena. Všechna hodnocení pravděpodobnosti popsaná v tomto pokynu předpokládají normalitu, pokud není uvedeno jinak.

POZNÁMKA Variabilita mezi jednotkami přidružená heterogenitě (heterosubclugeneitě) a změnám vyvolaným nestabilitou by neměla mít normální distribuci a může vyústit v asymetrická rozdělení.

- c) Hodnoty uváděné jako „indikační“, „informativní“ nebo jinak uváděné jako nezahrnuté do prohlášení metrologické návaznosti a nejistoty měření, se považují jako nevhodné pro metrologické aplikace jako je validace metod, kalibrace, nebo přidělení hodnot dalším materiálům.
- d) V rámci tohoto pokynu se používá pro vyhodnocení nejistoty měření zákon šíření nejistoty. Mohou se též použít i další metody vyhodnocování nejistoty měření a v některých případech je to tak nutné učinit. Další náměty pro tyto otázky lze nalézt v GUM a jeho doplňcích (viz odkazy <sup>[5]</sup> a <sup>[6]</sup>).

## 5 Conventions

In this Guide, the following conventions are used.

- a) A measurand is specified in such a way that there exists a unique 'true value'.
- b) All statistical methods used in this Guide are based on the following assumptions.
  - 1) The certified value is the best estimate of the true value of the property of the CRM.
  - 2) All variation, be it associated with the material (i.e. homogeneity and stability) or the measurement process, is random and follows a normal probability distribution or can be transformed to a normal distribution. All probability assessment described in this Guide assumes normality unless otherwise stated.

NOTE Variation between units associated with heterosubclugeneity and changes due to instability might not be normally distributed and can result in asymmetric distributions.

- c) Values given as "indicative", "informative", or otherwise identified as not being covered by statements of metrological traceability and measurement uncertainty, are considered to be unfit for use in metrological applications such as method validation, calibration, or the assignment of values to other materials.
- d) Throughout this Guide, the law of propagation of uncertainty is used for the evaluation of measurement uncertainty. Other methods of evaluating measurement uncertainty may also be applied, and in some cases, it is necessary to do so. Further guidance on these matters is given in the GUM and its supplements (see References <sup>[5]</sup> and <sup>[6]</sup>).

## 6 Přehled výroby referenčního materiálu

### 6.1 Obecně

Výroba a distribuce referenčního materiálu vyžaduje pečlivé plánování před tím, než začne jakékoliv uskutečňování projektu. Následující položky jsou stručným přehledem kroků, ze kterých sestává výroba referenčního materiálu s následujícím popisem hlavních otázek jednotlivých plánovaných kroků. Detailní návod pro hodnocení homogenity, stability a provedení charakterizace podávají kapitoly 7, 8 a 9 v uvedeném pořadí.

**POZNÁMKA** V tomto pokynu se *certifikačním projektem* rozumí výroba jakéhokoli referenčního materiálu, bez ohledu na to, zda vyhovuje definici *certifikovaného referenčního materiálu*.

### 6.2 Shrnutí návrhu projektu

Výroba referenčního materiálu sestává z následujících kroků:

- a) definice RM, tj. matrice, vlastností, které budou charakterizovány a jejich úrovní, zamýšleného použití materiálu a (pro CRM) požadované úrovně nejistoty,
- b) návrhu vzorkovacího postupu pro získání daného materiálu,
- c) návrhu postupu zpracování referenčního materiálu,
- d) výběru měřicích metod vhodných pro testování homogenity a stability,
- e) návrhu charakterizace referenčního materiálu,
- f) výběru vhodných metod pro charakterizaci, testování homogenity a studie stability,
- g) zvážení metrologické návaznosti u každé měřené vlastnosti, u které je prohlášení o metrologické návaznosti vyžadováno a to zejména pro certifikované referenční materiály,
- h) posouzení homogenity,
- i) posouzení stability
- j) posouzení komutability (pokud je vyžadováno)
- k) charakterizace referenčního materiálu,
- l) pro CRM pak spojení výsledků testování homogenity, testování stability a charakterizace (pokud jsou vyžadovány) k získání úplného odhadu nejistoty měření,
- m) přípravu certifikátu nebo informačního listu produktu a podle okolností i zprávy o výrobě a/nebo certifikaci,
- n) definování podmínek přepravy a skladování,

## 6 An 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. The following subclauses provide a brief overview of the steps involved in the production of a reference material followed by a description of the main issues involved in planning each step. Detailed guidance on homogeneity assessment, stability assessment and characterization is given in Clauses 7, 8 and 9, respectively.

**NOTE** In this Guide, the term *certification project* is understood to apply to the production of any reference material, whether or not it meets the definition of *certified reference material*.

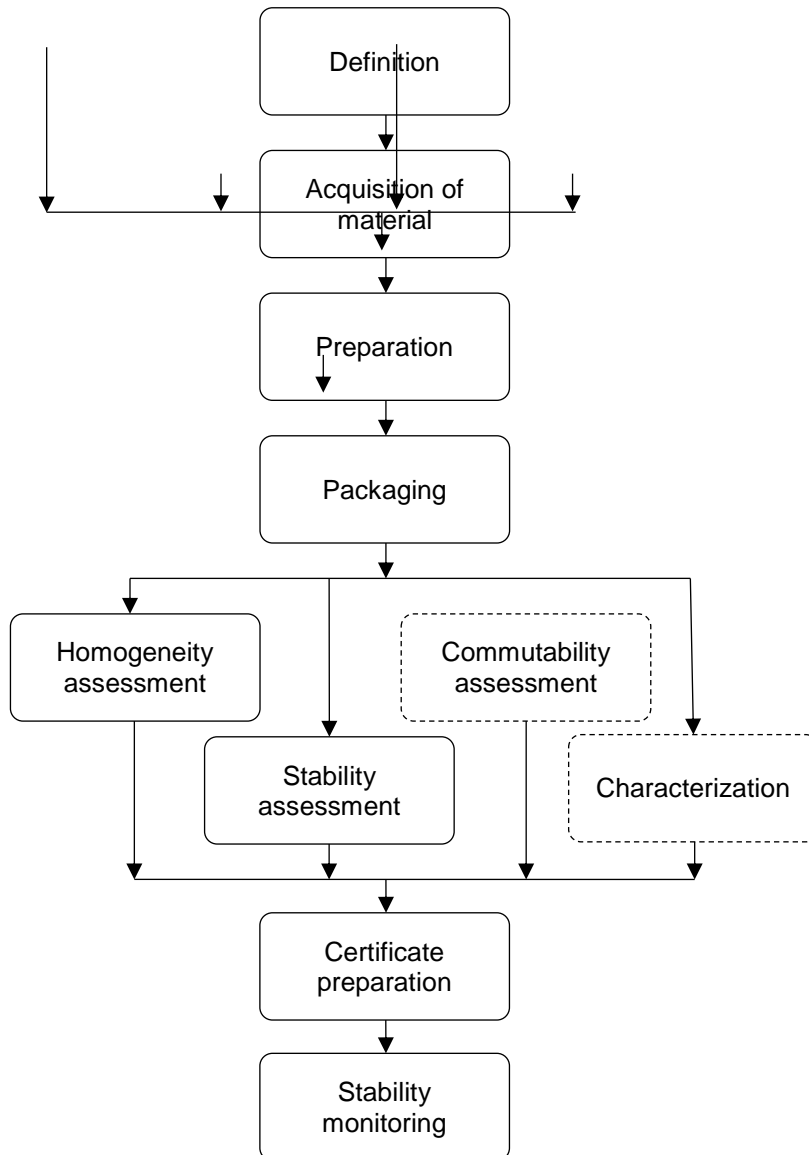
### 6.2 Summary of project design

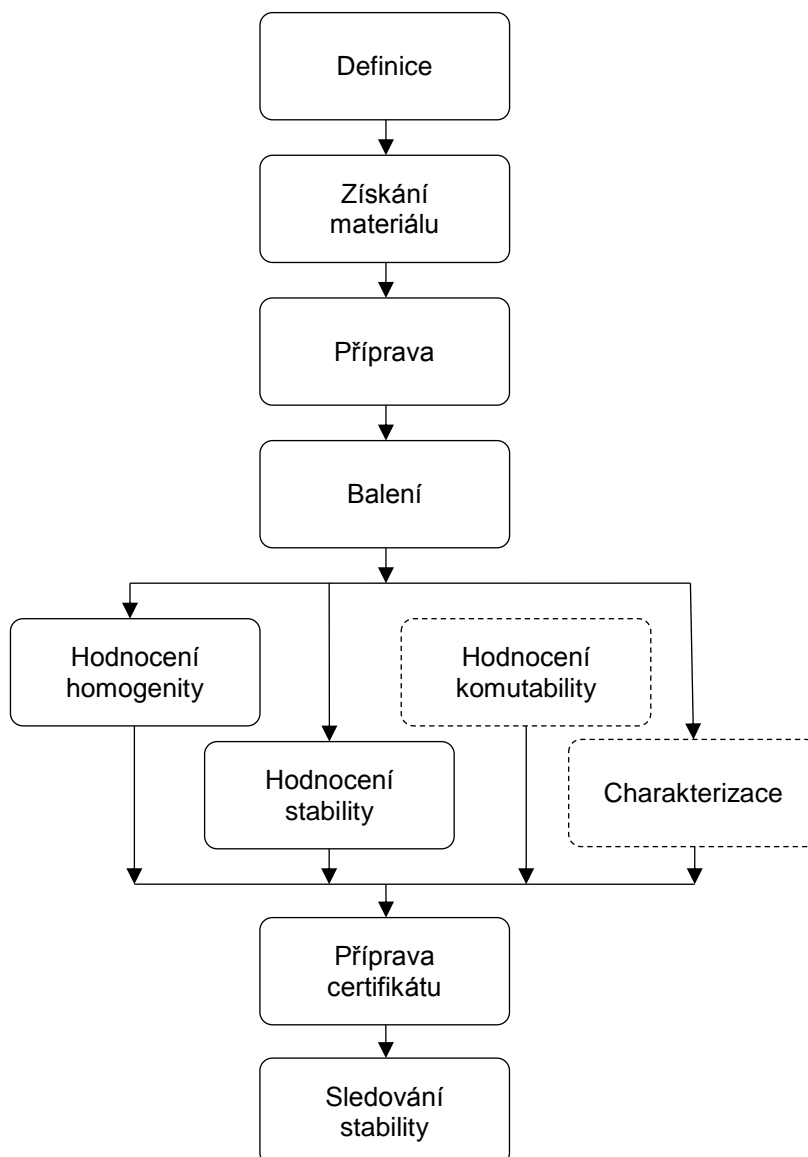
The production of a reference material involves the following steps:

- a) definition 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 level of uncertainty desired;
- b) design of a sampling procedure for the sourcing of the material;
- c) design of a reference material processing procedure;
- d) selection of measurement methods appropriate for homogeneity and stability testing;
- e) design of the characterization of the reference material;
- f) choice of suitable methods for characterization, homogeneity testing and stability studies;
- g) consideration of metrological traceability for each measured property, particularly for CRMs, for which a statement of metrological traceability is required;
- h) assessment of homogeneity;
- i) assessment of stability;
- j) assessment of commutability (if required);
- k) characterization of the reference material;
- l) for CRMs, combination of the results from homogeneity testing, stability testing and characterization (as required) to provide a full evaluation of the measurement uncertainty;
- m) preparation of a certificate or product information sheet and, if appropriate, a report on the production and/or certification;
- n) definition of storage and transportation conditions;

o) monitorování stability po skončení výroby.  
 Hlavní etapy schematicky znázorňuje obrázek 1.

o) post-production monitoring of stability.  
 The main stages are shown schematically in Figure 1.





POZNÁMKA 1 Obrázek poskytuje schematický nástin hlavních kroků přípravy a uchování referenčního materiálu. Položky s čárkovaným obrysem nejsou ve všech případech potřeba.

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

POZNÁMKA 2 'Balení' v tomto schématu zahrnuje rozdělení do jednotlivých jednotek a do vhodných obalů pro distribuci.

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

POZNÁMKA 3 'Příprava certifikátu' v tomto schématu představuje přípravu všech typů dokumentace, která by měla doprovázet referenční materiál, včetně certifikátu, informačního listu produktu, certifikační zprávy atd.

NOTE 3 In this diagram, 'Certificate preparation' includes all types of documentation that could be provided with a reference material, including a certificate, product information sheet, certification report, etc.

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

**Obrázek 1 — Schematický náčrt projektu referenčního materiálu**

### 6.3 Získání výchozího materiálu

Prvním úkolem projektu výroby RM je získání dostatečného množství výchozího materiálu

### 6.3 Acquisition of starting material

The first task in an RM production project is the acquisition of a sufficient amount of starting material(s)

(výchozích materiálů) požadovaných vlastností. O přípravě materiálu s určitými vlastnostmi pojednává stručně odstavec 6.5. Množství potřebného materiálu se řídí následujícím:

- počtem jednotek RM požadovaným pro distribuci v průběhu očekávané doby života RM,
- počtem jednotek potřebných pro studii homogenity,
- počtem jednotek potřebných pro studii stability,
- počtem jednotek potřebných pro charakterizaci kandidátského RM,
- počtem jednotek požadovaných pro sledování stability v průběhu očekávané životnosti materiálu,
- plánovaná velikost každé jednotky RM, která musí postačovat přinejmenším pro jedno měření,
- potřebě studie proveditelnosti,
- fakultativně jednotky navíc pro nepředvídané skutečnosti jako jsou např. následné studie vyvolané reakcí na otázky zákazníků, recertifikace v budoucnosti vyžádaná výraznou změnou podmínek skladování nebo rozšíření počtu certifikovaných vlastností.

Počet jednotek kandidátského RM, které jsou požadovány k distribuci je často, přinejmenším zčásti, komerční otázkou a měl by se pečlivě zvážit před uvedením souboru do provozu a náběhu zpracování materiálu. Kromě toho očekávaná dlouhodobá stabilita materiálu při skladování může ovlivnit množství suroviny, která může být připravena k použití. U méně stabilních materiálů může být rozumné omezit počet vyráběných jednotek, aby se zabránilo plýtvání způsobenému nevyhnutelnou degradací v průběhu času.

#### 6.4 Studie proveditelnosti

Studie proveditelnosti jsou krátké studie mající řešit obavy o proveditelnost výroby a charakterizaci dostatečně homogenního a stabilního RM: Například otázky jakými jsou nejlepší cesta přípravy RM, nebo zajištění dostatečné stability daného materiálu se mohou řešit pomocí studií proveditelnosti malého rozsahu na samém začátku projektu.<sup>[7][8]</sup>

Tam, kde se předpokládá, že charakterizace se bude provádět pomocí mezilaboratorní studie, může studie proveditelnosti identifikovat možné zdroje chyb a umožnit účastníkům podílejícím se na charakterizaci optimalizovat své vybavení a postupy.

**POZNÁMKA** Při studii proveditelnosti zaměřené na testování a zlepšování schopností účastníků při mezilaboratorním charakterizačním experimentu (viz kapitola 9), použití materiálu odlišujícího se od kandidátského RM může zabránit zbytečnému vychýlení výsledků účastníků vyvolanému předchozí znalostí kandidátského RM.

with the desired properties. The preparation of material with particular properties is considered briefly in 6.5. The amount of material needed is dictated 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;
- 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 at least sufficient for one measurement;
- the need for a feasibility study;
- 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 a candidate 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 raw material that can usefully be produced. It may be prudent 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.<sup>[7][8]</sup>

Where characterization is expected to be performed through the use of 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 aiming at testing and improving 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 Příprava referenčního materiálu

Příprava může zahrnovat celou řadu procesů včetně například:

- syntézu, výrobu nebo přípravu syntetického referenčního materiálu,
- sušení, lyofilizaci, mletí a/nebo filtraci u přírodních materiálů,
- přidavek stabilizačních činidel,
- homogenizaci před balením.

Použití příslušných postupů závisí na daném materiálu a obvykle vyžaduje odborné posouzení.

## 6.6 Hodnocení homogenity

Homogenita je důležitým požadavkem pro všechny referenční materiály a skládá se z homogenity uvnitř jednotek a homogenity mezi jednotkami. Homogenita mezi jednotkami je důležitá pro zajištění, že každá jednotka RM je nositelem stejné hodnoty pro každou vlastnost. Homogenita uvnitř jednotky je důležitá tam, kdy uživatelé materiálu mohou odebírat podvzorky k měření. O detailech hodnocení homogenity pojednává kapitola 7.

## 6.7 Hodnocení stability

Referenční materiály musí vykazovat dostatečnou stabilitu pro jejich zamýšlené použití, aby se koncový uživatel mohl spolehnout na přidělenou hodnotu v jakémkoli okamžiku doby platnosti certifikátu. Typicky je třeba zvažovat stabilitu v průběhu dlouhodobého skladování, za podmínek přepravy a (tam, kde to připadá v úvahu) i pro podmínky skladování v laboratoři uživatele. To může zahrnovat zvážení stability po otevření, pokud je dovoleno další použití. O detailech hodnocení stability pojednává kapitola 8.

## 6.8 Výběr postupů měření

Každý krok projektu výroby referenčního materiálu, který vyžaduje měření, může využívat odlišné měřicí postupy, protože každé z hodnocení stability, hodnocení homogenity i charakterizace mají odlišné požadavky na měřicí metody. Například charakterizace (přidělení hodnoty) vyžaduje postupy vykazující minimální vychýlení a nízkou nejistotu, testování homogenity vyžaduje v první řadě co nejlépe dostupnou opakovatelnost; klasické studie stability vyžadují metody s dobrou precizností v průběhu doby v rámci stejné laboratoře. Výběrem měřících postupů pro studie homogenity, testování stability a charakterizaci se zabývají kapitoly 7, 8 a 9 v uvedeném pořadí.

## 6.9 Metrologická návaznost

Metrologická návaznost je klíčem k zajištění srovnatelnosti výsledků měření v čase a mezi různými místy, včetně těch, které se používají k charakterizaci referenčních materiálů. Podle definice jsou

## 6.5 Reference material preparation

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

- synthesis, manufacture or formulation of a synthetic reference material;
- drying, lyophilisation, milling, and/or filtration for natural materials;
- addition of stabilizing agents;
- homogenization prior to packaging.

The particular 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 may be taken for measurement by users of the material. Clause 7 gives detailed guidance on homogeneity assessment.

## 6.7 Stability assessment

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 transportation conditions and, (where applicable) the storage conditions at the RM user's laboratory. This may include consideration of stability after opening, if re-use is permitted. Clause 8 provides detailed guidance on stability assessment.

## 6.8 Choice of measurement procedures

Each step in a reference material production project which requires measurements may use different measurement procedures, because stability testing, homogeneity testing and characterization each place different demands on the measurement method. For example, characterization (value assignment) generally requires minimally biased procedures with low uncertainty; homogeneity testing primarily requires the best available repeatability; and classical stability studies typically require methods that show good precision over time within the same laboratory. The choice of measurement procedures for homogeneity studies, stability testing and characterization is considered in Clauses 7, 8 and 9, respectively.

## 6.9 Metrological traceability

Metrological traceability is key in ensuring the comparability of measurement results over time and between locations, including those used to characterize reference materials. By definition, CRMs



certifikované referenční materiály doprovázeny prohlášením o metrologické návaznosti pro každou certifikovanou hodnotu vlastnosti. Vhodný výběr „stanovených referencí“, ke kterým je metrologická návaznost hodnot vlastností stanovena, je stěžejní otázkou koncipování certifikovaných referenčních materiálů, protože certifikované referenční materiály se především používají k zajištění toho, aby výsledky měření byly návazné. O zajištění metrologické návaznosti pojednává podrobně odstavec 9.2.

### 6.10 Charakterizace a vyhodnocení nejistoty

Charakterizace znamená přidělení hodnot odpovídajícím vlastnostem RM. O charakterizaci RM pojednává kapitola 9. U certifikovaných referenčních materiálů jsou certifikované hodnoty doprovázeny prohlášením o nejistotě měření. Vyhodnocením nejistoty se zabývá kapitola 10.

### 6.11 Hodnocení komutability

Komutabilita referenčního materiálu se vztahuje ke schopnosti daného RM, charakterizovaného jedním měřicím postupem (obvykle referenčním postupem) fungovat jako kalibrátor nebo materiál pro řízení kvality i pro druhý měřicí či zkušební postup, který se používá u rutinních zkušebních materiálů. Toto je zvláště důležité tam, kde různé měřicí postupy mohou reagovat velmi odlišným způsobem na různé typy zkušebních materiálů. Hodnocení komutability se nevyžaduje u všech referenčních materiálů, ale je požadováno u některých významných tříd referenčních materiálů.

POZNÁMKA V současném stanovisku REMCO k hodnocení komutability <sup>[9]</sup> se uvádí, že:

Výrobce referenčních materiálů by měl provést hodnocení komutability když

- a) zamýšlené použití vyžaduje komutabilitu kalibračních materiálů nebo materiálů pro řízení kvality a
- b) výrobce referenčních materiálů zaručuje, že daný materiál je vhodný pro zamýšlené použití.

POZNÁMKA 1 Prokázání komutability se obvykle vyžaduje, když zamýšlené použití zahrnuje kalibraci nebo řízení kvality u biologických měření a obvykle se nevyžaduje, když biologická měření zamýšlené použití nezahrnuje a o postupu je známo, že je dostatečně specifický vůči měřené veličině v matici referenčního materiálu a zamýšlených rutinních vzorcích.

POZNÁMKA 2 Obvykle není nutné potvrzovat komutabilitu, když referenční materiál a jeho provenience jsou získány ze stejných zdrojů a zpracovány stejným způsobem jako vzorky, které budou zkoušeny pro zákazníky, například pro maticové referenční materiály.

### 6.12 Otázky přepravy

Téměř všechny referenční materiály se musí přepravovat na místo použití. Prostředky a podmínky přepravy referenčního materiálu z výroby souvisejí

are accompanied by a statement of metrological traceability for each certified property value. The proper choice of the “stated references” to which metrological traceability of the property values is established, is a major design issue 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 assignment of values to the relevant properties of an RM. Characterization of an RM is described in Clause 9. For CRMs, certified values are accompanied by a statement of measurement uncertainty; the evaluation of uncertainty 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 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 some important classes of RM.

NOTE Current REMCO information on commutability assessment <sup>[9]</sup> states that:

“A reference material producer should conduct an assessment of commutability where

- a) the intended use requires commutability of calibration or quality control materials, and
- b) the reference material producer warrants that the material is fit for intended use.

NOTE 1 Demonstration of commutability is usually required when the intended use includes calibration or quality control in biological measurement, and is not usually 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 reference material and the intended routine samples.

NOTE 2 It is not usually necessary to establish commutability when the reference material and its origin are obtained from sources and handled the same as samples that would be tested for customers, for example, matrix reference materials.”

### 6.12 Transport issues

Nearly all RMs have to be transported to the location of use. The means and conditions of transport of an RM after production are relevant to the need for

s požadavky na studii stability (viz kapitola 8) a je tedy potřebné zvažovat podmínky přepravy již v počátečním stadiu projektu.

**POZNÁMKA 1** Národní a/nebo mezinárodní přepravní předpisy mohou omezovat možnosti pro přepravu, zakázat přepravu některých materiálů, nebo vyžadují zvláštní obaly nebo opatření pro bezpečnost nebo z jiných důvodů.

**POZNÁMKA 2** Doba potřebná pro úřední postupy, jako například celní nebo jiné hraniční kontroly může zvýšit lhůty přepravy u některých destinací.

### 6.13 Monitorování stability

Většina referenčních materiálů se skladuje po delší dobu v prostorách výrobce referenčních materiálů nebo u distributorů. Jelikož hodnocení stability nemůže obvykle předvídat všechny změny, ke kterým může dojít, je obvykle potřebné, jako součást managementu rizika spojeného s možnou nestabilitou, monitorovat hodnoty vlastností materiálů uchovávaných po delší dobu. Protože požadavky na monitorování závisí z části na znalostech získaných při hodnocení stability, je návod na monitorování stability součástí odstavce 8.10.

### 6.14 Referenční materiály vyráběné v opakovaných šaržích

Potřebnost experimentální studie některých charakteristik (zejména homogenity, stability a komutability) se může snížit tam, kde materiál se vyrábí opakovaním výrobního postupu jako u předchozích materiálů podle stanovené procedury. Spoléhání se na předchozí zkušenosti je rozumné, pokud:

- a) proces výroby šarží RM není nezměněn žádným způsobem, který by mohl mít nepříznivý vliv na jeho koncové použití,
- b) materiály použité ve výrobě RM se nezměnily žádným způsobem, který by mohl mít nepříznivý vliv na koncové použití RM,
- c) materiály z předchozí výroby vyrobené stejným procesem nevykazují žádné nedostatky, které by se mohly přičíst procesu výroby, ať už během rutinního monitorování nebo od uživatelů a
- d) požadavky na materiál jsou pravidelně přezkoumávány se zřetelem na zamýšlené použití materiálu v době přezkoumávání, aby se zajistilo, že proces výroby zůstává vhodný pro daný účel.

Měla by se kontrolovat stabilní výkonnost procesu výroby, například srovnáváním hodnot vlastností vzorků z následujících šarží za podmínek opakovatelnosti.

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 may limit the options for transport, prohibit the transport of some materials, or require specific packaging or precautions for safety or other reasons.

**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 Stability monitoring

Most reference materials are stored for extended periods at the RM producer's premises or by distributors. Since stability assessment cannot usually anticipate all changes that may 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.

### 6.14 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 of previous materials produced following an established procedure. Reliance on prior experience is reasonable so long as:

- a) the process for producing batches of the RM does not change in any way that may adversely affect the end use;
- b) the materials used in production of the RM do not change in any way that may adversely affect the end use;
- c) materials previously produced by the same process show no failures attributable to the production process, either during routine monitoring or by users; and
- 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 Hodnocení homogenity

### 7.1 Úvodní část

Většina referenčních materiálů se připravuje v šaržích ‚jednotek‘ (např. lahvíček, lékovek nebo zkušebních vzorků). Je důležité, aby všechny distribuované jednotky byly stejné v rámci uváděné nejistoty každé certifikované hodnoty vlastnosti a pokud nejsou dodávány jako jednotky pro jedno použití, aby materiál v každé jednotce byl stejný. V souladu s tím Pokyn ISO 34 <sup>[2]</sup> vyžaduje hodnocení homogenity referenčního materiálu.

Homogenita se může vztahovat buď k variabilitě hodnoty vlastnosti mezi jednotlivými jednotkami materiálu, nebo k variabilitě uvnitř každé jednotky. Je tedy nutné vždy hodnotit variabilitu mezi jednotkami. Tam, kde zamýšlené použití dovoluje použití části jednotky – například malého podílu pevného nebo kapalného materiálu, nebo malé části povrchu – je obvykle dodatečně nutné buď hodnotit variabilitu uvnitř materiálu (heterogenitu uvnitř jednotky) nebo poskytnout instrukce pro použití, které budou mít dopad heterogenity uvnitř jednotky pod kontrolou. Tyto instrukce mohou obsahovat například promíchání vzorku a u zrnitých materiálů minimální odběr vzorku, protože heterogenita uvnitř jednotky se přímo odráží v minimální velikosti podvzorku, který je reprezentativní pro celou jednotku.

Hodnocení homogenity může zahrnovat použití předchozích dokladů (včetně přechozích experimentálních důkazů) o homogenitě materiálu, provedení experimentální studie homogenity na kandidátském referenčním materiálu nebo obou zdrojů. Ve většině případů je nutné provést experimentální studii. Výjimky představují, například šarže vysoce homogenního materiálu, jako je roztok, u kterého předchozí experimentální studie prokázaly, že balení a skladování neovlivňují homogenitu, nebo výroba materiálů pro které každá jednotka referenčního materiálu má přidělenou separátní hodnotu. V odstavci 7.2 jsou uvedeny další podrobnosti o okolnostech a typech materiálu vyžadujících experimentální studii.

Studie homogenity tvoří důležitou část řízení kvality a prokazování kvality výroby referenčních materiálů. Výsledky experimentální studie homogenity se obvykle užívají k výpočtu jedné ze složek nejistoty certifikačního modelu (viz kapitola 10). Velikost této složky nejistoty se může měnit podstatným způsobem ve srovnání s dalšími složkami nejistoty, v závislosti na charakteru referenčního materiálu a certifikované vlastnosti.

K provedení studie homogenity se vybírá podskupina jednotek z dané šarže, obvykle 10 až 30, hodnoty vlastností se u každé jednotky měří za použití vhodných měřících metod a výsledky se hodnotí pomocí vhodných statistických metod, aby se získaly

## 7 Assessment of homogeneity

### 7.1 Preamble

Most RMs are prepared as batches of ‘units’ (e.g. bottles, vials or test pieces). It is important that all distributed units are the same within the stated uncertainty for each certified property value and, unless sold as single-use units, that the material within each unit is uniform. ISO Guide 34 <sup>[2]</sup> accordingly requires assessment of the homogeneity of a reference material.

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 usually additionally 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 intake, 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 may include the use of prior evidence (including prior experimental evidence) of the homogeneity of the material, performing an experimental homogeneity study on the candidate reference material, or 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 reference material unit has a separate assigned value. Subclause 7.2 gives further details about the circumstances and types of material requiring experimental study.

Homogeneity studies are often an important part of quality control and assurance for RM production. The results of an experimental homogeneity study is 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 to 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, property values are measured for each unit using a suitable method of measurement and the results are assessed using appropriate statistical methods to obtain

informace o například variabilitě mezi jednotkami a variabilitě uvnitř jednotek daného materiálu.

Pro získání spolehlivých výsledku je důležité

- vybrat reprezentativní skupinu jednotek,
- vybrat vhodnou měřicí metodu o postačující opakovatelnosti a selektivitě,
- provést měření za vhodných podmínek podle patřičného experimentálního uspořádání a
- provést statistickou analýzu za použití platných statistických metod.

V následujících odstavcích je uveden návod pro každý z těchto kroků.

Historicky, studie homogenity testují statisticky významné rozdíly mezi jednotkami ve srovnání s precizností měření, aby se rozhodlo, že materiál je homogenní či nikoliv. Tento pokyn nezastává tento přístup, spíše klade důraz na rozhodování, zda směrodatná odchylka mezi jednotkami je dostatečně malá pro zamýšlené koncové použití. Statistické testy významnosti se však mohou používat ve výrobě RM, například pro rozhodování, zda je potřeba další zpracování ke snížení heterogenity až na nevýznamnou úroveň ve srovnání s rutinní precizností měření. Pokud se používají statistické testy, musí experiment zaměřený na homogenitu být schopen detekovat závažnou heterogenitu a vyžaduje dostatečně vyhovující kombinaci preciznosti měřicího postupu, počtu jednotek RM a počtu replikátů u jednotky. Výpočty statistické síly testu (7.4.2) mohou napomoci k zajištění, že test je dostatečně účinný.

## 7.2 Potřeba experimentální studie homogenity

Materiály přírodního původu nebo ty s komplexními maticemi jako jsou potraviny, zeminy, rudy a slitiny jsou ve složení typicky heterogenní. I když velikost rozdílů mezi jednotkami může být někdy malá nebo zanedbatelná, mohou rozdíly mezi jednotkami převyšovat nejistotu přidělené hodnoty vlastnosti (charakterizaci). Referenční materiály připravené z takových heterogenních materiálů by měly tudíž být podrobeny experimentální studii homogenity, aby se posoudila jejich heterogenita.

Očekává se, že referenční materiály připravované jako čisté látky nebo roztoky čistých látek (pokud jsou certifikovány na čistotu, ne na nečistoty) vykazují vysoký stupeň homogenity. Tyto materiály však mohou vykazovat určitou heterogenitu, například vlivem gradientu hustoty, vypařování rozpouštědla během zpracování a plnění nebo u kovů obsahujících proměnná množství okludovaných plynů. Dále se očekává, že certifikované hodnoty takových materiálů mají velmi malé nejistoty, což činí potenciálně významnými i malá množství heterogenity. I v případech, kdy by se očekávala perfektní homogenita, měla by se tato prověřit. Prověření může znamenat kompletní studii homogenity nebo jinou

information on, for example, between-unit variability and within-unit variability of the material.

To obtain reliable results, it is important to

- select a representative subset of units,
- choose a suitable measurement method with sufficient repeatability and selectivity,
- make the measurements under suitable conditions following an appropriate experimental design, and
- conduct the statistical analysis using valid statistical methods.

The following paragraphs provide guidance on each of these steps.

Historically, homogeneity studies have tested for statistically significant between-unit differences compared with measurement precision in order to decide whether a material is homogeneous or not. This approach is not taken in the present Guide; rather, emphasis is placed on deciding whether the between-unit standard deviation is sufficiently small for the intended end use. Statistical tests of significance may, however, be of use in RM production, for example in order to decide whether further processing is required to reduce heterogeneity to insignificance compared to routine measurement precision. If statistical tests are used, however, the homogeneity experiment shall 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 assignment of the property value (characterization). RMs prepared from such heterogeneous materials should therefore be subjected to an experimental homogeneity study to assess the heterogeneity.

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, evaporation of solvent during processing or filling, 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 perfect homogeneity might be expected, homogeneity should be verified. Verification may include a complete homogeneity study or other check (for example, a

kontrolu (například prověření konzistence bodů tání jednotek čistého organického materiálu).

Experimentální studie homogenity není nezbytná v následujících případech“

- když materiál pochází z opakované výroby stejným postupem jako u předchozího materiálu, u kterého se experimentálně prokázalo, že je dostatečně homogenní nebo
- tam, kde postup výroby byl validován, čímž se prokázalo, že produkuje soustavně homogenní šarže materiálu.

Příklady materiálů, které jsou někdy vyráběny tímto způsobem, jsou etanolové kalibrační roztoky nebo prvkové kalibrační roztoky připravené navažováním a důkladným promícháním.

Tam, kde se prokázání homogenity opírá o validovaný proces výroby, měly by se používat postupy řízení kvality k potvrzení konzistentního chodu procesu výroby. Mezi takové postupy patří například používání regulačních diagramů rozsahu či směrodatné odchylky pro rozsah nebo směrodatnou odchylku malého počtu měřených jednotek, nebo používání kritérií pro rozsah hodnot nalezených při každé charakterizaci.

**POZNÁMKA** Odstavec 6.12 podává dodatečný návod pro spoléhání na zkušenosti získané z předchozích výrobních šarží.

### 7.3 Vlastnosti, které mají být předmětem studia

Tam, kde má být homogenita určena na základě experimentu, je obvykle potřeba stanovit homogenitu pro každou vlastnost, která je předmětem zájmu, tedy každou vlastnost s certifikovanou nebo jinak přidělenou hodnotou.

Homogenita všech předmětných vlastností se může hodnotit zkoumáním menšího počtu vybraných vlastností když

- existuje dostatečně vysoká statistická korelace mezi konkrétními hodnotami vlastností daného typu materiálu umožňující efektivní předpověď jedné hodnoty vlastnosti z jedné či několika dalších a
- může se prokázat, že konkrétní skupiny vlastností dostatečně těsně souvisí (například svojí přítomností v konkrétní složce směsi) takže měření jedné vlastnosti takové skupiny poskytuje důkaz homogenity dalších vlastností ve stejné skupině.

Je nezbytné, aby jakákoliv podskupina vlastností, která byla zvolena jako reprezentant homogenity pro větší soubor vlastností byla vhodně vybrána na základě známých chemických nebo fyzikálních vztahů. Například současný výskyt prvků v minerálních součástech RM by byl podporou předpokladu, že RM vykazuje podobný stupeň homogenity i pro prvky, které nebyly začleněny do výběru.

V případech, kdy se homogenita hodnotí experimentálně za použití podskupiny předmětných

check on melting point consistency between units of a pure organic material).

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.

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

**NOTE** Subclause 6.12 provides additional guidance regarding reliance on experience gained from previous production batches.

### 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 with a certified value or other assigned value.

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

- there is sufficiently high statistical correlation between particular property values in this type of material to allow useful prediction of one property value from one or more others, and
- it can be shown that particular groups of properties are sufficiently closely associated (for example, 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 on the basis of 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

vlastností, měl by se získat dodatečný doklad o homogenitě vlastností, které nebyly hodnoceny experimentálně. Takový důkaz bude postačující, pokud prokáže, že nejistota přidružená k heterogenitě vlastností, které nebyly studovány experimentálně, není podceněna.

**POZNÁMKA** Takový důkaz se může získat například z literatury, týkající se dotyčného materiálu, nebo prostřednictvím studie stability nebo charakterizace materiálu.

## 7.4 Statisticky platná schémata vzorkování

### 7.4.1 Minimální počet jednotek pro studii homogeneity

#### 7.4.1.1 Studie homogeneity kvantitativních vlastností

Studie homogeneity kvantitativních vlastností se obecně zaměřují na získání informací o variabilitě pocházející z heterogeneity a na možné (třeba nelineární) trendy mající původ ve zpracování. Pro dosažení těchto cílů je potřeba, aby počet položek postačoval pro přiměřený odhad rozptylu mezi jednotkami a aby odebrané položky poskytl jasný obraz o možných trendech, které se vyskytují.

Vycházejíce ze současné praxe, přijatelný odhad mezijednotkové variability lze získat s devíti či více stupni volnosti. Při plánu jednoduché studie homogeneity a znárodněnému blokovému plánu (viz níže) to odpovídá výběru minimálně 10 jednotek. Tam, kde se použije pro měření hierarchický plán, ve kterém podskupina jednotek se měří v každé z několika sérií, měly by se začlenit do studie další jednotky, aby se zachovaly požadované stupně volnosti. V odstavci 7.6.4 se pojednává podrobněji o potřebném počtu dalších jednotek.

Trendy pocházející ze zpracování se často projevují jako prvopočáteční trend následovaný stabilním výstupem, jako trend, který se projevuje později v procesu nebo jako kombinace obou. Tam, kde se vyskytnou tyto projevy, je často možné získat homogenní materiál vyloučením postižených jednotek z počátku nebo konce dané série. Prozkoumání pouze 10 jednotek nemusí však poskytnout dostatek informací o počátku trendů v koncích zdlouhavého průběhu zpracování. V současné nejlepší praxi se tedy zvyšuje počet zkoumaných jednotek s tím, jak se zvyšuje celkový počet vyráběných jednotek  $N_{\text{prod}}$ . Typická doporučení jsou mezi  $\sqrt[3]{N_{\text{prod}}}$  a  $3 \times \sqrt[3]{N_{\text{prod}}}$ . Pokud zvážíme společně požadavek na stupně volnosti uvedený výše, pak je minimální počet jednotek  $N_{\text{min}}$  pro testování homogeneity materiálů charakterizovaných na kvantitativní vlastnosti dán jako

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

**PŘÍKLAD** Výrobce připravuje 3 000 jednotek kandidátského RM a má v úmyslu provést jednoduchou

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 instance, from literature relevant to the material in question, or 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 items should be sufficient to give a reasonable estimate of the between unit variance, and sufficient items taken 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 (see below), 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 in order to maintain the required degrees of freedom. Subclause 7.6.4 gives further details about the additional number of units required.

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 may 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  $3 \times \sqrt[3]{N_{\text{prod}}}$ . Taken together with the degrees of freedom requirement above, this leads to a recommended minimum number of units  $N_{\text{min}}$  for homogeneity testing of materials characterized for a quantitative property given by

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

**EXAMPLE** A producer prepares 3 000 units of a candidate RM and intends to undertake a simple homogeneity

studii homogenity v jediné sérii. Pak  $N_{\text{prod}} = 3\,000$  a  $N_{\text{min}} = \max(10, \sqrt[3]{3\,000})$ , což je 14,4. Tato studie tudíž vyžaduje pro studii homogenity 15 jednotek materiálu.

#### 7.4.1.2 Studie homogenity pro kvalitativní (“jmenovité”) vlastnosti

U referenčních materiálů certifikovaných na kvalitativní, či „jmenovité“ vlastnosti počet jednotek vybraných pro studii homogenity by měl představovat soubor zvolený na základě návodu pro statistické přejímky srovnáváním tak, jak jej popisuje ISO 2859 (všechny části) [8] nebo podobného návodu.

Plány vzorkování pro přejímku srovnáváním vedou k velmi vysokým počtům pro zkoumání, jestliže nízké podíly vad jsou detekovány pouze na základě vzorkování. Například k detekci 1 % vad s vysokou jistotou vycházející z 1 a více zjištěných vad musí být zkoumáno okolo 300 náhodně vybraných jednotek (toto číslo vychází ze síly statického výpočtu pro sílu testu 95 % pro 1 % vad). To je často nerealistické pro typické velikosti šarží referenčních materiálů. U kvalitativních materiálů se tudíž experimentální testování homogenity pravděpodobně omezí na kontrolu hrubé heterogenity (například větší než 10 % vad) s dalšími informacemi o původu materiálu, jeho zpracování, atd., které jsou podkladem pro prohlášení o homogenitě.

Jestliže se certifikovaná kvalitativní vlastnost jednotlivě ověřuje u každé jednotky takového materiálu, není nutné provádět u takového materiálu další testování homogenity.

#### 7.4.1.3 Malé výrobní šarže

Některé referenční materiály se vyrábějí ve velmi malých šaržích, například čítajících 50 nebo méně jednotek. Sekundární kalibrační standardy (etalony) plynů vyráběné národními metrologickými instituty jsou jedním z častých příkladů. Pro takové malé výrobní šarže, minimální počet jednotek specifikovaných v 7.4.1.1 obvykle představuje velmi velký podíl z jednotek, které jsou k dispozici. Tam, kde velikost šarže se pohybuje pod 100 jednotkami, hodnocení homogenity by se mělo opírat o hodnocení pomocí 3 a výše jednotek nebo 10 % z velikosti šarže, náhodně vybíraných z dané šarže. Opakování by mělo být tak vysoké, jak je prakticky proveditelné pro získání co nejlépe dostupné síly testování počtu použitých jednotek. Analýza síly (7.4.2) může pomoci při zvažování požadovaných úrovní opakování, například u třech jednotek, čtyři pozorování na jednotku poskytují přibližně stejnou sílu detekce odchylek mezi jednotkami u normálního rozdělení jako duplikátní pozorování 10 jednotek, i když s mnohem nižší pravděpodobností detekce nahodilých vadných jednotek v jinak homogenní šarži.

### 7.4.2 Použití analýzy statistické síly

Analýza statistické síly testů může napomoci při volbě vhodného počtu jednotek a opakování u studie

study in a single run. Then  $N_{\text{prod}} = 3\,000$  and  $N_{\text{min}} = \max(10, \sqrt[3]{3\,000})$ , 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 reference materials 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 ISO 2859 (all parts) [8] or similar guidance.

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

If a certified 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 reference materials are produced in very small batches, for example of 50 or fewer units. Secondary gas calibration standards produced by national metrology institutes (NMIs) are one common example. For such small production batches, the minimum number of units specified in 7.4.1.1 usually represents a very large fraction of the available units. Where the batch size is below 100 units, homogeneity should be assessed on the larger of three 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 for the number of units used. Power analysis (7.4.2) may assist in considering desirable replication levels; for example with three units, four observations per unit gives approximately the same power of detecting normally distributed between-unit deviations as observing 10 units in duplicate, though with much lower probability of detecting occasional defective units in an otherwise homogeneous batch.

### 7.4.2 Use of statistical power analysis

Statistical power analysis may assist in choosing a suitable number of units and replicates for the

homogeneity. Analýza síly se snaží řídit pravděpodobnost nezjištění určitého stupně heterogenity dané v navrhovaném statistickém testu pro významnou heterogenitu. Analýza síly je speciálním tématem, ale je ve zvyšující se míře dostupná v softwarech, některých dostupných i zdarma. Nejběžnějším příkladem analýzy síly v tomto kontextu je výpočet počtu položek testu a opakování za předpokladu, že se používá analýza rozptylu jednoduchého třídění (viz 7.7.3) k testování významných vlivů mezi jednotkami.

Ačkoliv analýza statické síly může být užitečná při porovnávání různých navrhovaných strategií, je třeba při jejím použití postupovat s opatrností. Zvláště volba počtu opakování a počet jednotek velmi silně závisí na předpokládaném rozdělení hypotetických rozdílů mezi vzorky. Například předpoklad normálního rozdělení pro (pravé) střední hodnoty jednotlivých jednotek, který je běžnou výchozí hodnotou u softwarů pro výpočet síly, vede k vysokému navrhovanému počtu replikátů a malému počtu jednotek. To je špatná volba, jestliže nejpravděpodobnější typ heterogenity je malý podíl neshodných jednotek v převážně homogenní populaci. Pokud chybí správná informace o pravděpodobném rozdělení jednotlivých (pravých) středních hodnot jednotek, je tudíž analýza síly nejvíce užitečná jako pomůcka pro volbu počtu opakování po tom, co byl určen navrhovaný počet jednotek.

**POZNÁMKA 1** Při použití analýzy síly typickou volbou významnosti testu a určené síly je 95 % a 80 % v uvedeném pořadí, i když je třeba poznamenat, že toto poskytuje podstatně větší sílu než minimální počet jednotek a opakování uvažovaných v 7.4.1.1.

**POZNÁMKA 2** Podrobnosti o některých jednoduchých výpočtech síly vhodných pro studie referenčních materiálů poskytuje odkaz [58].

### 7.4.3 Strategie vzorkování u studie homogenity

Schéma vzorkování pro výběr jednotek (položek, lahvíček) u studie homogenity je typicky buď prostým náhodným vzorkováním, náhodným stratifikovaným vzorkováním nebo systematickým vzorkováním. U všech jsou obecné postupy při aplikaci na vzorkování referenčního materiálu pro testování homogenity následující:

- Při prostém náhodném vzorkování se odebírá náhodný vzorek z dané šarže.
- Při stratifikovaném náhodném vzorkování se šarže typicky rozdělí na několik segmentů stejné velikosti (obvykle podle výrobního sledu nebo sledu balení nebo někdy polohy) a odebírá se náhodně stejný počet jednotek (často jediná) z každého takového segmentu.
- Při systematickém vzorkování se stanoví interval  $n_{\text{sys}}$  mezi vzorkovanými jednotkami a náhodným počátečním bodem  $n_1$  mezi 1 a  $n_{\text{sys}}$ , a odebírají se jednotky při  $n_1, n_1 + n_{\text{sys}}, n_1 + 2n_{\text{sys}}, \dots$  atd.  $n_1$  a

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 number of test items and replicates on the assumption that one-way analysis of variance (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 when using power analysis. 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 units, a common default in power calculation software, leads to a high proposed number of replicates and a small number of units. This is a poor choice if the most likely pattern of heterogeneity is a small proportion of discrepant 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 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 it should be noted that this provides substantially greater power than the minimum number of units and replicates considered at 7.4.1.1.

**NOTE 2** Details of some simple power calculations appropriate for reference material studies are given in Reference [58].

### 7.4.3 Sampling strategy for homogeneity study

The sampling scheme used to pick the units (items, bottles) for a homogeneity study is typically one of simple random sampling, random stratified sampling or systematic sampling. The general procedures for each, as applied to reference material sampling for homogeneity testing, are as follows.

- Simple random sampling takes a random sample from the batch.
- Stratified random sampling typically divides the batch into a number of segments of equal size (usually by production or packaging sequence or, sometimes, location) and takes an equal number of units (often one) at random from each such segment.
- Systematic sampling sets an interval  $n_{\text{sys}}$  between sampled units and a random start point  $n_1$  between 1 and  $n_{\text{sys}}$ , and takes units at  $n_1, n_1 + n_{\text{sys}}, n_1 + 2n_{\text{sys}}, \dots$  etc.  $n_1$  and  $n_{\text{sys}}$  are chosen



$n_{\text{sys}}$  se volí tak, aby výsledkem bylo úplné pokrytí dané výrobní šarže a požadovaný počet jednotek.

Náhodné vzorky by se měly určit použitím software na generaci náhodných čísel nebo pomocí tabulek náhodných čísel. Manuální libovolný výběr není ekvivalentem náhodného výběru.

Stratifikované náhodné vzorkování se doporučuje pro většinu situací, protože zaručuje, že jednotky odebírané pro studii homogenity jsou přibližně rovnoměrně rozděleny v celé šarži a zabraňuje koincidenci mezi místem vzorkování a jakýmkoli cyklickými vlivy při zpracování. Systematické vzorkování se může použít tam, kde je malé riziko přehlédnutí opakujících se vlivů nebo trendů v dané šarži. Prosté náhodné vzorkování je nevhodnější tam, kde není známo nějaké uspořádání nebo pro malé šarže, ve kterých se použije velký podíl jednotek pro testování homogenity.

Schéma vzorkování by mělo brát v úvahu potenciální slabiny metody přípravy a skladování vzorků (včetně například trendů při zpracování a plnění a možného usazování při skladování před rozdělováním) a tak umožňovat kritické zkoumání vyrobené šarže.

Jednotky by se normálně měly očíslovat před vzorkováním, nebo číslovat v pořadí zpracování po vzorkování aby bylo možné provést následně analýzu trendů.

Je možno použít alternativní schémata vzorkování tam, kde se může prokázat, že výsledné odhady rozptylu nejsou vychýlené za použití zvoleného vzorkovacího schématu.

## 7.5 Výběr a provádění měřicího postupu u studie homogenity

### 7.5.1 Výběr měřicího postupu

Měřicí postup, vybraný k provádění měření u testování homogenity, by se měl v první řadě volit podle dobré preciznosti v průběhu očekávané doby trvání každé série měření (tedy dobré směrodatné odchytky opakovatelnosti  $s_r$ ) a pokud se berou jednotky náhodně z několika sérií, pak dobré preciznosti mezi sériemi (malá  $s_1$ ). Směrodatná odchytky měření by měla být malá ve srovnání s očekávanou nejistotou pro každou měřenou vlastnost, v ideálním případě by měla být směrodatná odchytky opakovatelnosti u postupu testování homogenity menší než jedna třetina požadované standardní nejistoty. Tedy je-li požadovaná standardní nejistota hodnoty vlastnosti  $u_d$ , směrodatná odchytky opakovatelnosti by měla ideálně vyhovovat vztahu

$$s_r \leq u_d/3 \quad (2)$$

Pokud měřicí metoda zvolená pro testování homogenity vykazuje směrodatnou odchytku opakovatelnosti větší než jedna třetina požadované standardní nejistoty, je třeba začlenit opakování měření všude tam, kde je to možné, aby

to give complete coverage of the production batch and the desired number of units.

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

Stratified random sampling is recommended for most situations, since this guarantees that the 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 may be applied when there is little risk of overlooking repetitive effects or trends in the batch. Simple random sampling is most appropriate when there is no known ordering, or for small batches in which a large proportion of the units is used for homogeneity testing.

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.

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

Alternative sampling schemes may 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 measurement procedure for homogeneity study

### 7.5.1 Choice of measurement procedure

The measurement procedures chosen to make measurements for homogeneity testing should be chosen primarily for good precision during the expected duration of each measurement run (that is, a good repeatability standard deviation  $s_r$ ) and, if units are to be randomized amongst several runs, good between-run precision (small  $s_1$ ). The standard deviation for measurements should be small compared to the expected uncertainty for the value for each property measured; ideally, the repeatability standard deviation for the homogeneity testing procedure should be less than one third of the desired standard uncertainty. That is, if the desired standard uncertainty for the property value is  $u_d$ , the repeatability standard deviation for the procedure should ideally comply with

$$s_r \leq u_d/3 \quad (2)$$

If the measurement method chosen for homogeneity testing has a repeatability standard deviation greater than one third of the desired standard uncertainty, sufficient replicates should be included wherever possible so that

$$\frac{s_r}{\sqrt{n_{al}}} \leq \frac{u_d}{3} \quad (3)$$

kde  $n_{al}$  je počet pozorování u každého z  $n$  alikvotů odebraných z každé jednotky  $k$  měření.

Tam, kde tento požadavek na preciznost nemůže být v praxi splněn, například když jednotky RM neumožňují provádět opakovaná měření, mělo by se vynakládat co největší úsilí na zajištění, aby  $s_r/\sqrt{n_{al}}$  nepřekročilo  $u_d$  a mělo by se zvážit zvýšení počtu jednotek.

**POZNÁMKA 1** Zvýšení počtu jednotek není tak efektivní jako zvýšení počtu pozorování u jednotky nebo zlepšení preciznosti měření, ledaže by předmětem zájmu bylo zjištění vysoce neshodných jednotek, vyskytujících se řídkce. Pro volbu vhodného počtu jednotek a opakování je možné použít výpočty statistické síly (viz 7.4.2), aby tak vzniklo experimentální schéma, které by zajistilo splnění požadavků na preciznost podobným způsobem, jako je to u výrazů (2) a (3) uvedených výše.

**POZNÁMKA 2** Předmětem další úvahy při výběru metody pro testování je množství a komplexnost přípravy vzorku, potřebných pro získání každého vzorku ve formě pro měření. Zkušební postupy vyžadující malou nebo žádnou přípravu vzorku (takzvané nedestruktivní metody) mohou vykazovat lepší preciznost než metody vyžadující extenzivní přípravu vzorků.

## 7.5.2 Provádění měření u studií homogeneity

V odstavci 7.6.1 se uvádí, že je důležité tam, kde je to možné, porovnávat disperzi výsledků u různých jednotek RM s disperzí za podmínek opakovatelnosti. „Podmínky opakovatelnosti“ zahrnují opakování kompletního postupu měření. Je tedy důležité, aby tam, kde je to možné, na každý podvzorek nebo alikvot odebraný z každé jednotky byly aplikovány kompletní příprava vzorku („transformace“) a finální měřicí krok (viz obr. 2). Jestliže se tak nečiní, každá změna u přípravy vzorku při opakovaných pozorováních bude zvyšovat výraz pro heterogenitu mezi jednotkami (obrázek 3) a bude mít za následek nadhodnocený odhad nejistoty přiřazené k heterogenitě.

$$\frac{s_r}{\sqrt{n_{al}}} \leq \frac{u_d}{3} \quad (3)$$

where  $n_{al}$  is the number of observations on each of  $n$  aliquots taken from each unit for measurement.

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  $s_r/\sqrt{n_{al}}$  does not exceed the desired standard uncertainty  $u_d$  and consideration should be given to increasing the number of units.

**NOTE 1** Increasing the number of units is not as effective as increasing the number of observations per unit or improving measurement precision unless the focus is on finding infrequent, highly discrepant units. Statistical power calculations (see 7.4.2) may be of use in choosing a suitable number of units and replicates to give an experimental design that provides similar assurance to meet the precision requirements in Formulae (2) or (3) above.

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

## 7.5.2 Conduct of measurements for homogeneity studies

Subclause 7.6.1 notes that it is important, where possible, to compare the dispersion of results for different RM units with the dispersion under repeatability conditions. “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 or aliquot taken from each 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.

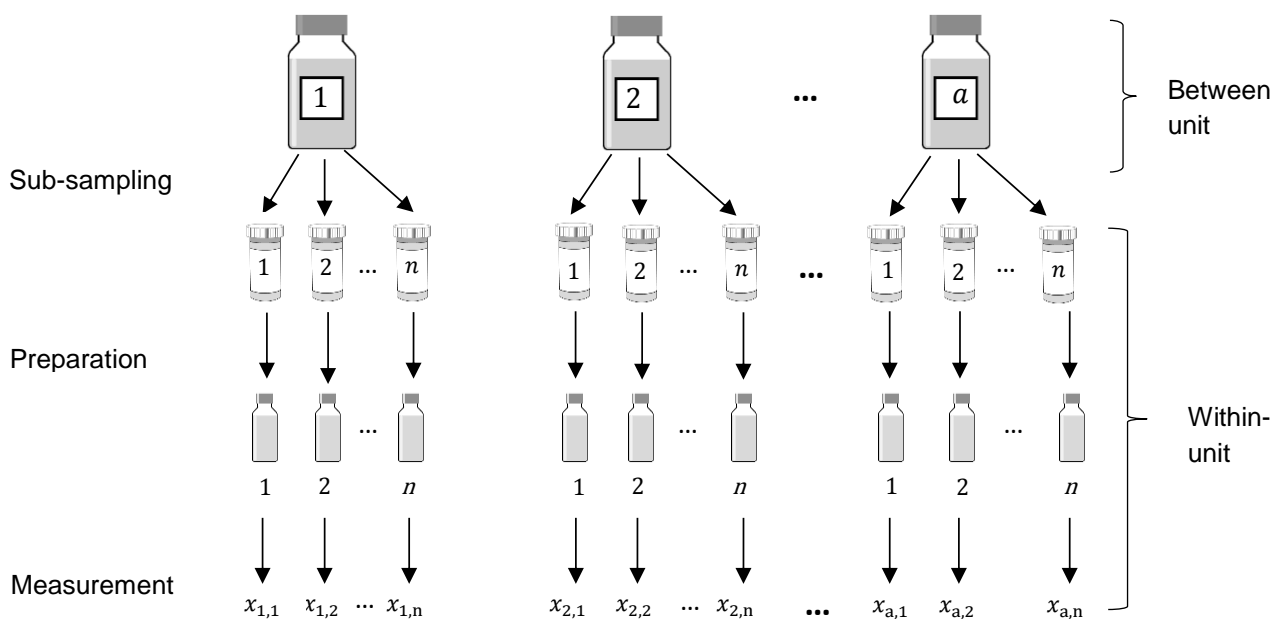
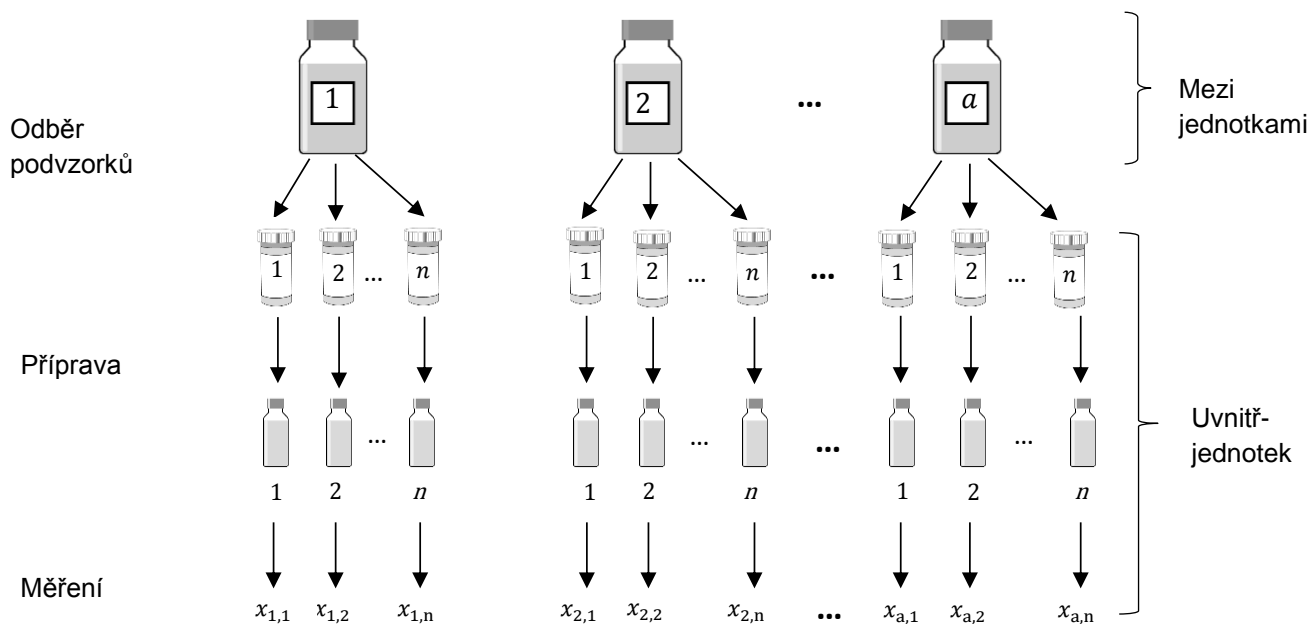


Figure 2 — Layout of a between-unit homogeneity study



Obrázek 2 — Uspořádání studie homogenity mezi jednotkami

Obrázek 2 znázorňuje ideální případ, ve kterém je možný odběr podvzorků z jednotek referenčního materiálu a je možno odebrat (n) zkoušených vzorků z každého vzorku šarže a upravit jednotlivě (transformovat) pro měření. V tomto případě rozptyl mezi jednotkami zahrnuje změny náležející odběru podvzorků, přípravě a měření. Pro záměr získat nevychýlený odhad směrodatné odchylky mezi jednotkami je toto ideální situace protože žádný z těchto vlivů nepřispívá k odhadu rozptylu mezi jednotkami.

In Figure 2, an ideal case is shown in which subsampling of reference material units is possible and multiple (n) test portions have been taken from each sample of the batch and individually prepared (transformed) for measurement. In this case, the variance within 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.

Obrázek 3 znázorňuje uspořádání pro případ, kdy odběr podvzorků z jednotlivých položek není možný nebo se zrovna neprovedl, z každé jednotky referenčního materiálu se odebral jeden podvzorek a provádí se  $n$  měření připraveného zkoušeného vzorku. V tomto uspořádání jsou vlivy přípravy vzorku a odběru podvzorku zahrnuty ve zjištěném rozptylu mezi jednotkami vypočteném z takového experimentu. Vypočtený rozptyl uvnitř jednotek zahrnuje pouze opakovatelnost měření. Tento postup poskytuje vyšší odhad rozptylu mezi jednotkami a následně příliš vysoký odhad heterogenity mezi jednotkami.

Z toho vyplývá, že tam, kde je to možné, měly by se studie homogenity provádět tak, že všechny kroky měřicího postupu se realizují na každém podvzorku.

In Figure 3, a layout is shown for the case where subsampling of the items is impossible or just not carried out, only one subsample from each of the  $a$  reference material 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 sub-sampling 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.

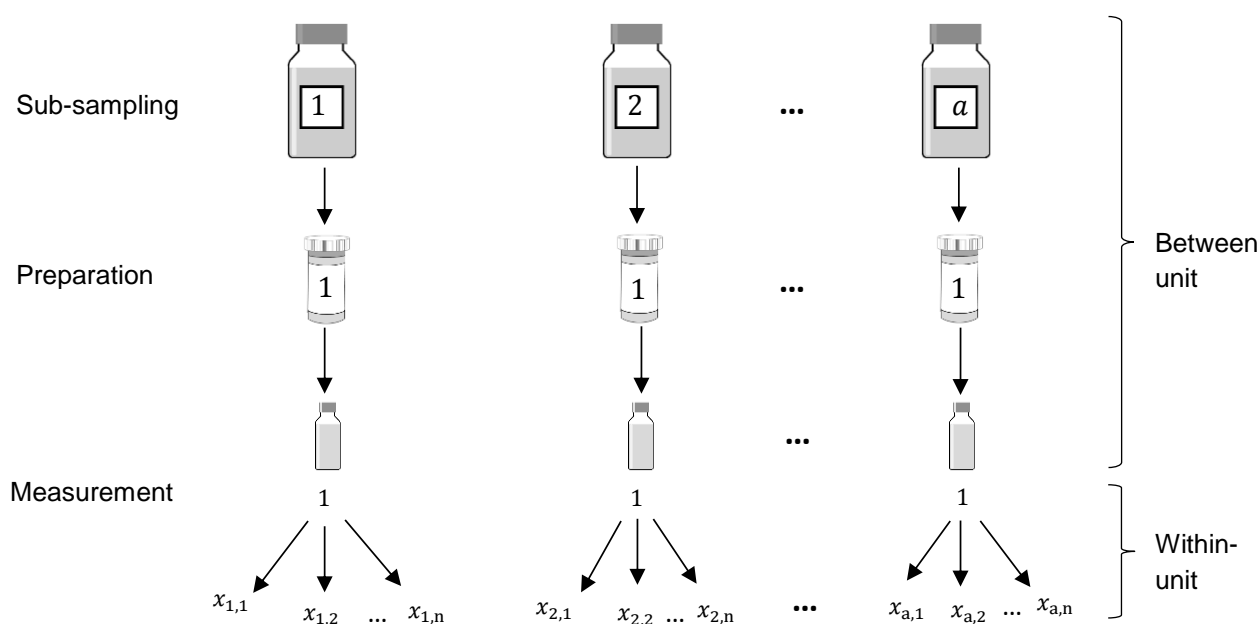
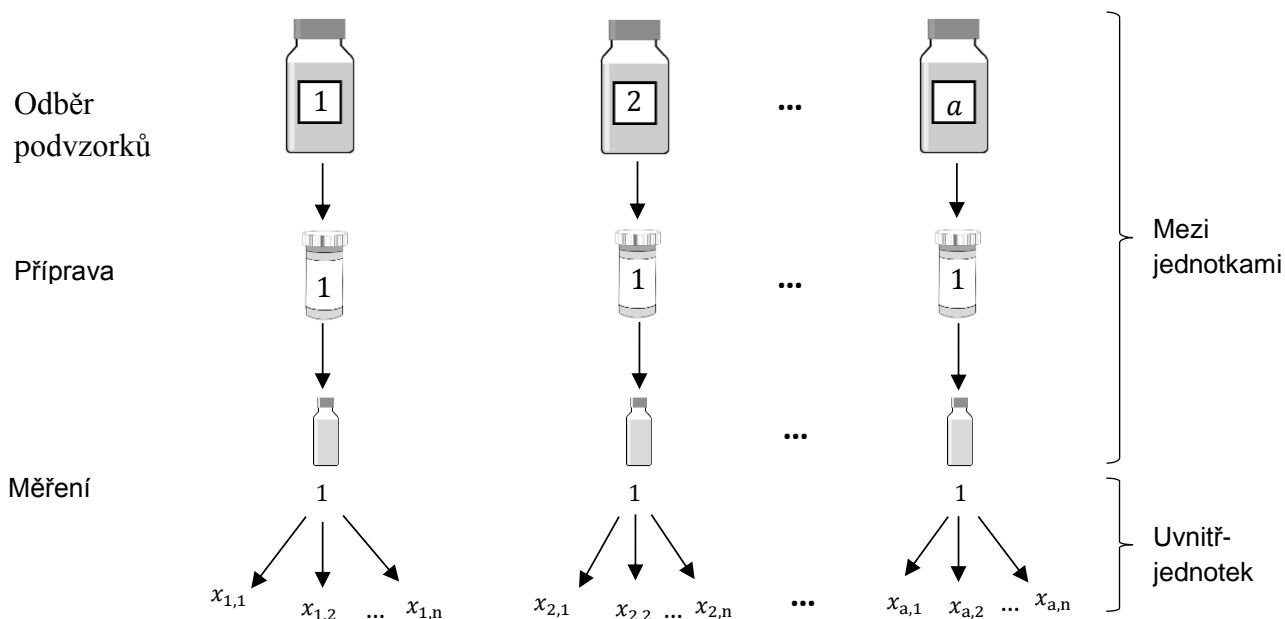


Figure 3 — Alternative layout of a between-unit homogeneity study



Obrázek 3 — Alternativní uspořádání studie homogenity mezi jednotkami

## 7.6 Návrhy studie homogenity

### 7.6.1 Cíl studie homogenity

Studie homogenity pomocí statistického vyhodnocení porovnává disperzi naměřených hodnot u několika jednotek RM s precizností metody měření s cílem zjistit směrodatnou odchylku mezi jednotkami, kterou lze pak použít pro výpočet nejistoty přiřazené heterogenitě. To obvykle představuje provedení stejného počtu opakovaných měření u každé jednotky RM, jak to znázorňuje obr. 2. K získání co nejlepšího odhadu hodnoty členu mezi jednotkami, disperze hodnot získaných z jednotek referenčního materiálu by se měla porovnávat s disperzí výsledků získaných, tam kde je to možné, za podmínek opakovatelnosti (viz odkaz <sup>[11]</sup>). Toho lze dosáhnout pomocí

- prováděním všech měření homogenity dané vlastnosti v jedné měřicí sérii, nebo
- použitím experimentálního uspořádání, které dovolí separátní odhad rozptylů v sérii, mezi sériemi a mezi jednotkami. Mezi nejpříhodnějšími experimentálními návrhy pro tento účel patří jednoduché znáhodněné blokové návrhy a vyvážené hierarchické návrhy.

Typické experimenty za použití těchto odlišujících se návrhů jsou popsány podrobněji níže. Uvažované tři návrhy znázorňuje obrázek 4.

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

## 7.6 Homogeneity study designs

### 7.6.1 Objective of a homogeneity study

A homogeneity study uses statistical evaluation to compare the dispersion of observations on several units of the RM to the precision of the measurement method in order 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 possible estimate of the between-unit term, the dispersion of values obtained for reference material units in a homogeneity study should be compared with the dispersion of results under repeatability conditions wherever possible (see Reference <sup>[11]</sup>). This can be achieved by

- conducting all homogeneity measurements for a given property in a single measurement run, or
- 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.

Typical experiments using these different designs are described in more detail below. The three designs considered are illustrated in Figure 4.

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

**Klíč**

- a ...Jednoduchý znáhodněný návrh, jedna série se všemi jednotkami měřená duplicitně v náhodném pořadí.
- b ...Znáhodněný blokový návrh. Každá série zahrnuje jedno měření každé jednotky, každá série je separátně znáhodňována.
- c ...Hierarchický návrh. Čtyři jednotky jsou náhodně přiděleny do každé ze tří sérií a u každé jsou provedena duplikátní měření, opět v náhodném pořadí. Všimněme si, že znáhodnění nezaručuje oddělení replikátů, zvláště u malého počtu jednotek v tomto schematickém příkladu hierarchického návrhu.

Všechna znázornění se týkají návrhů pro duplikátní měření po 12 jednotek RM **XXX**

d) 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**

- 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. Note that randomization does not guarantee separation of replicates, especially for the small number of units in this schematic example of a nested design

All illustrations show designs for duplicate observations on 12 RM units **designated a to l**

**Figure 4 — Designs for homogeneity testing****Obrázek 4 — Experimentální návrhy testování homogenity****7.6.2 Základní návrh testování homogenity – měření v jedné sérii**

Tam, kde měřicí systém je stabilní po požadovanou dobu provádění měření, všechna měření se mohou provést v jediné měřicí sérii. Měření by se měla provádět takovým způsobem, aby

- trendy měřicího systému nebyly špatně posuzovány jako rozdíly mezi jednotkami RM a
- trendy měření mohly být, pokud je to potřeba, separovány od trendů v šarži vzorků.

Toho lze nejlépe dosáhnout měřením replikátů vzorků používaných ve studii homogenity ve zcela náhodném pořadí. Obrázek 4 uvádí ilustrativní příklad takového návrhu.

Výrobci mohou též použít systematické uspořádání opakovaných měření tam, kde se může prokázat, že

- důvodně očekávané typy chyby měření, včetně trendů a skokových změn odezvy v sérii nebudou příčinou znatelného vychýlení odhadu rozptylu mezi jednotkami a
- následná kontrola a postupy analýzy dat budou schopny v průběhu studie homogenity rozlišit trendy zpracování (včetně balení) od trendů měření.

**POZNÁMKA 1** Plně znáhodněné pořadí nezaručuje, že jednotlivé replikáty od té samé jednotky RM nebudou spolu sousedit. Občasný výskyt sousedících replikátů je běžný a přijatelný u plně znáhodněného návrhu.

**7.6.2 The basic homogeneity test design – measurement in a single run**

Where the measurement system is stable for the duration of the observations required, all observations may be acquired in a single measurement run. Measurements should be carried out in such a way that

- trends in the measurement system are not misinterpreted as differences between RM units, and
- 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.

Producers may 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, and
- subsequent inspection and data analysis procedures are able to distinguish processing (including packaging) trends from any measurement trends during the homogeneity study.

**NOTE 1** A fully 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.

Náhodně zvolené pořadí může shodou okolností vést k jednotkám seřazeným téměř v pořadí, jak byly zpracovány nebo k tomu, že všechny replikáty té samé jednotky RM jsou při pořadí měření těsně u sebe. Navrhované znáhodněné pořadí, které představuje zřejmou silnou korelaci s produkční sekvencí nebo těsné seskupení replikátů u většiny nebo všech jednotek by se mělo odmítnout a zvolit nové náhodné pořadí.

**POZNÁMKA 2** Následující příklad demonstruje systematické seřazení vzestupné a sestupné <sup>[12]</sup>, které nezapříčiní znatelné vychýlení členu mezi jednotkami při výskytu jednoduchého lineárního trendu v čase v přístrojové odezvě.

**PŘÍKLAD** Předpokládejme, že se pro studii homogenity používá 10 vzorků s trojí replikací. Možné systematické schéma pro provedení měření vypadá takto:

Opakování č. 1: 1 – 3 – 5 – 7 – 9 – 2 – 4 – 6 – 8 – 10

Opakování č. 2: 10 – 9 – 8 – 7 – 6 – 5 – 4 – 3 – 2 – 1

Opakování č. 3: 2 – 4 – 6 – 8 – 10 – 1 – 3 – 5 – 7 – 9

### 7.6.3 Znáhodněný blokový návrh

U znáhodněného blokového návrhu pro  $n_0$  replikátů z každé z  $a$  jednotek referenčního materiálu, nejjednodušší znáhodněný blokový návrh představuje  $n_0$  měřících sérií, kde každá jednotka je měřena jednou v náhodném pořadí. Série by měly být znáhodňovány jednotlivě. Obrázek 4b) představuje ilustrativní příklad takového návrhu. Dvoufaktorová analýza rozptylu bez opakování nebo jiné metody (viz níže) se mohou použít pro odhad směrodatné odchylky uvnitř jednotek a mezi jednotkami nezávisle na vlivu série.

Znáhodněný blokový návrh je vhodný tam, kde nezbytný počet replikátů u všech jednotek se nemůže zahrnout do jedné série kvůli časovým nebo instrumentálním omezením. Je též platný tam, kde měřicí proces je náchylný k driftu, který se projevuje na konci dlouhotrvající série. Za těchto okolností lze získat lepší preciznost (měřenou jako směrodatnou odchylku uvnitř série) použitím několika kratších sérií v znáhodněném blokovém návrhu, než by se mohlo dosáhnout v rámci jedné série.

**POZNÁMKA 1** Při klasické statistické analýze (použití ANOVA s výpočtem ze sumy čtverců), stupně volnosti pro směrodatnou odchylku mezi jednotkami zůstávají pro jednoduchý znáhodněný blokový návrh rovny  $a-1$ , ale residuální stupně volnosti, spojené s rozptylem uvnitř jednotek, se snižují o jednu pro každou přídatnou sérii.

**POZNÁMKA 2** Při jednoduchém znáhodněném blokovém návrhu, jak je popsáno výše, vypočtená směrodatná odchylka uvnitř jednotek zahrnuje jak variabilitu měření, tak variabilitu náležející heterogenitě uvnitř jednotky.

A randomly chosen ordering can, by chance, lead to 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 units should be rejected and a new random ordering selected.

**NOTE 2** A systematic forward-reverse ordering of observations <sup>[12]</sup> that does not cause appreciable bias in the between-unit term in the presence of a simple linear trend in instrument response with time is illustrated in the following example.

**EXAMPLE** Suppose 10 samples are used for a homogeneity study, with 3 replicates. A possible systematic scheme for conducting the measurements reads as follows:

Replicate No. 1: 1 – 3 – 5 – 7 – 9 – 2 – 4 – 6 – 8 – 10

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  $a$  units of a reference material, the simplest randomized block design involves  $n_0$  measurement runs and each unit is measured once in random order. Runs should be randomized individually. Figure 4 b) shows an illustrative example of such a design. Two-way analysis of variance without replication, or other methods (see below), 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 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, but 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 above, the calculated within-unit standard deviation includes both variation due to the measurement and variation due to any within-unit heterogeneity.

### 7.6.4 Vyvážený hierarchický návrh

Při vyváženém hierarchickém návrhu pro  $a$  jednotek se provádí  $n_r$  sérií měření, každá zahrnuje stejný počet  $a_w$  jednotek, které se měří s  $n_0$  replikáty v rámci stejné série. Každá série se znáhodňuje odděleně. Na obrázku 4c) je znázorněn ilustrativní příklad.

Hierarchické návrhy snižují počet stupňů volnosti spojených se směrodatnou odchylkou mezi jednotkami. U těchto návrhů by se měl zvýšit počet studovaných jednotek, aby se zachoval dostatečný počet stupňů volnosti. To obvykle vyžaduje přidání minimálně jedné jednotky pro každou další sérii měření v rámci hierarchického návrhu. Kromě toho, pokud použitá statistická metoda se opírá o vyvážený návrh, měly by se přidat další jednotky RM, aby se zajistilo, že každá série měření obsahuje stejný počet jednotek RM.

**POZNÁMKA** Hierarchické návrhy zjednodušují analýzu při použití klasické ANOVA založené na sumě čtverců. Je výhodné držet stejný počet jednotek ve skupině, je-li analýza rozptylu prováděna manuálně nebo pomocí software (jako jsou aplikace v tabulkovém procesoru, které poskytují základní ANOVA funkce). Novější metody odhadu rozptylu (viz 7.7) jsou výpočetně složitější, ale jsou méně ovlivňovány nevyvážeností.

**PŘÍKLAD** Výrobce používá metodu testování homogenity umožňující až 12 měření, v jedné šarži připravuje 2 000 jednotek kandidátského RM. Vzorec (1) udává  $N_{\min} = 12,6$ , či 13 jednotek, pokud studie je završena v jediné sérii. Jelikož výrobce nemůže získat duplikátní pozorování u 13 jednotek v rámci jediné série, zvolil výrobce použití hierarchického návrhu experimentů ve třech sériích. Znamená to přidání dvou dalších sérií a počet jednotek je tudíž zvýšen o 2 na 15. Výrobce tudíž provádí měření 15 jednotek ve třech skupinách po pěti jednotkách.

### 7.6.5 Alternativní strategie

Ne vždy je možné uskutečnit jeden z experimentálních návrhů uvedených výše. Například jednotky referenčního materiálu nemusí umožňovat další dělení pro opakované analýzy, třeba proto, že množství k dispozici postačuje jen pro jedno měření, protože jednotky jsou nestálé po otevření, nebo protože jednotky jsou nedělitelné (například jednotlivé vzorky pro stanovení rázové houževnatosti metodou Charpy). Mezi další situace patří velmi krátké série měření (například tři měření v instrumentální sérii) nebo potřeba brát v úvahu další zdroje variability kromě vlivu v sérii.

Tam, kde jednotky nelze dělit, je nejběžnější strategií měřit každou jednotku jednou a porovnávat směrodatnou odchylku (nebo častěji rozptyl) těchto měření s nezávislým odhadem opakovatelnosti měření. Například rozptyl jednotlivých měření 10 jednotek se mohou porovnávat s rozptylem 10 pozorování pečlivě zhomogenizované přípravy stejného materiálu nebo s dobrým odhadem preciznosti vycházejícím z validačních studií. Rozdíl

### 7.6.4 Balanced nested design

In a balanced nested design for  $a$  units, a number of runs  $n_r$  are conducted, each including the same number  $a_w$  of 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 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 units per group if analysis of variance is to be performed manually or by software (such as spreadsheet applications that provides basic ANOVA functionality). More recent variance estimation methods (see 7.7) are computationally more complex but much less affected by imbalance.

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

### 7.6.5 Alternative strategies

It is not always possible to carry out one of the designs listed above. For example, reference material units may not permit subdivision for replicate analysis, perhaps because the amount of material provided is sufficient only for a single measurement, because units are insufficiently stable once opened, or because the units are indivisible (for example, individual Charpy test pieces). Other situations include very short measurement runs (for example, three observations per instrument run) or a need to allow for a further source of variation in addition to run effect.

Where 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 units might be compared to the variance of 10 observations on a carefully homogenized preparation of the same material, or to a good estimate of precision



v rozptylu je pak brán jako odhad rozptylu mezi jednotkami.

Tak kde je k dispozici jen délka série, která je příliš krátká, aby umožnila úplný znáhodněný blokový nebo hierarchický návrh experimentů, mohou používané strategie například zahrnovat:

- náhodné rozvržení replikátů v krátkých sériích s následnou statistickou analýzou jako u základního návrhu. To může vést ke zvýšení rozptylu uvnitř jednotek, protože vliv v sérii se zahrnuje do členu uvnitř skupiny.
- použití ‚vyváženého neúplného blokového‘ návrhu, kde pečlivě zvolená množina replikátů se získá u každé z mnoha krátkých sérií. Vyvážené neúplné blokové návrhy jsou mimo rozsah tohoto Pokynu, ale jsou popsány v odkazovaných textech zabývajících se návrhy experimentů.<sup>[11]</sup> Statistická analýza může používat klasické nebo i jiné metody.

Tam, kde je třeba mít pod kontrolou četné zdroje variability, se mohou použít další dostupné experimentální návrhy včetně (například) návrhů replikovaných latinských čtverců<sup>[12]</sup> nebo jiných strukturovaných návrhů experimentů.<sup>[11]</sup>

## 7.7 Vyhodnocení studie homogenity

### 7.7.1 Prvotní inspekce trendů v měření a odlehklých hodnot

#### 7.7.1.1 Trendy v měření

Experimentální data (včetně sledování řízení kvality) by se měla nejprve prozkoumat, zda se v nich neobjevují trendy měřicího procesu. Toho se nejefektivněji dosáhne vizuální inspekcí vnesení dat v experimentálním pořadí. Regresní analýza jako je lineární regrese s jednoduchým modelem přímky se může použít pro odhalení statisticky významného trendu, pokud je zřejmý přibližně lineární trend.

Trend v měření u vhodně znáhodněného experimentu homogenity bude obecně zvyšovat rozptyl uvnitř jednotek a potenciálně zastiňovat významné vlivy mezi jednotkami. Pokud se významný trend u výsledků měření potvrdí, je třeba přijmout opatření k minimalizaci jeho vlivu na následnou analýzu. Mezi typická opatření patří:

- Vyřazení postižené série měření (sérií měření) a opakování měření po odstranění příčin výskytu trendu.
- Použití korekce dat vycházející z vhodné vyhlazovací funkce (například lineární regrese) před provedením následné analýzy rozptylu. Přednostní korekce trendu snižuje residuální stupně volnosti a toto by se mělo brát v úvahu při následném odhadu rozptylu.
- Zahrnout trend do analýzy dat. Současný odhad trendu a odpovídajících členů rozptylu se může

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 may include, for example:

- random allocation of replicates amongst short runs, followed by statistical analysis as for the basic design. This may 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 Guide but are described in reference texts on experimental design.<sup>[11]</sup> Statistical analysis may use classical or other methods.

Where multiple sources of variation are to be controlled, other available designs may be used, including (for example) replicated Latin square designs<sup>[12]</sup> or other structured experimental designs.<sup>[11]</sup>

## 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 quality control 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, may 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 (for example, a linear regression) before conducting subsequent analysis of variance. Prior correction for a trend reduces the residual degrees of freedom and this should be taken into account in subsequent variance estimation.
- Include the trend in the data analysis. Simultaneous estimation of a trend and the relevant variance

dosáhnout použitím například modelování smíšených vlivů za použití odhadu pomocí metody omezené maximální věrohodnosti. Modely „smíšených vlivů“ odhadují současně rozptyly náhodných vlivů, zejména vlivu mezi jednotkami a koeficienty pevných vlivů jakým je směrnice a úsek trendu. Tento postup vhodně zohledňuje ztrátu stupňů volnosti, avšak vyžaduje speciální statistický software a příslušnou zkušenost ve statistice.

POZNÁMKA 1 Příklady takových alternativ lze nalézt v příloze C.

POZNÁMKA 2 Od výrobce se očekává, že určí kritéria kdy trend je významný. Při absenci jiných kritérií se při testování homogenity obvykle bere statistická významnost na 95% konfidenční úrovni jako indikace významného trendu v měření

POZNÁMKA 3 V některých případech, například u hierarchického nebo znáhodněného blokového návrhu experimentů se třemi nebo více sériemi, může vynechání jediné nezdařené série ponechat dostatek dat pro efektivní posouzení homogenity, aniž je třeba opakovat nezdařenou sérii.

#### 7.7.1.2 Odlehlé hodnoty v základní studii homogenity

V rámci studie homogenity se vytváří mnoho výsledků měření a technické problémy mohou způsobovat občasný výskyt odlehlých hodnot, jejichž příčina leží v instrumentaci nebo jiných přiřaditelných příčinách. Příčiny vzniku odlehlých hodnot by se měly zkoumat a měla by se uplatňovat patřičná nápravná opatření. Mezi nápravná opatření v takových případech patří, například:

- odstranění jednotlivého odlehlého pozorování nebo všech výsledků jednotky materiálu, u které se takové pozorování vyskytlo (viz Poznámka 2 níže),
- přeměření pochybné jednotky nebo podvzorku v krátké době,
- opakování kompletní série měření nebo celého experimentu, ve kterém se vyskytlo několik odlehlých výsledků.

Není vhodné vyloučit výsledky testování homogenity jednotky (certifikovaného) referenčního materiálu na základě toho, že vykazují extrémní střední hodnotu pro tuto jednotku RM.

POZNÁMKA 1 Odlehlé hodnoty uvnitř jednotky mohou mít příčinu v heterogenitě uvnitř jednotky. Pokud je heterogenita uvnitř jednotky pravděpodobná je rozumné provést další zkoumání před tím, než se jakýkoliv replikátní odlehlý výsledek vyloučí.

POZNÁMKA 2 Vynechání jednotlivé odlehlé hodnoty z jedné jednotky vyústí v nerovnovážnost. Tam, kde použita statistická metoda je založena na vyváženém návrhu experimentů, data této jednotky se mohou vynechat pod podmínkou, že neodlehlé hodnoty této jednotky jsou

terms can be achieved using, for example, mixed effects modelling using restricted maximum likelihood estimation. “Mixed effects” models simultaneously estimate variances for random effects, particularly the between-unit effect, and coefficients for fixed effects, such as slope and intercept of a trend. This procedure adjusts properly for any loss in degrees of freedom, though specialist statistical software and appropriate statistical expertise are required.

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

NOTE 2 The 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 testing.

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.

#### 7.7.1.2 Outliers in the basic homogeneity study

Many measurement results are generated as part of a homogeneity study and technical issues may result in occasional outlying results due to instrumental or other attributable causes. The cause of outliers should be investigated and appropriate corrective action should be applied. Corrective action in such a case may 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 below);
- 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.

It is not appropriate to reject homogeneity testing results for a (certified) reference material unit on the basis of an extreme mean value for that RM unit.

NOTE 1 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 2 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

v souladu se zbytkem datového souboru a že počet zbylých jednotek ve studii bude postačovat.

**POZNÁMKA 3** Větší studie homogenity (například studie o 20 nebo více jednotkách RM pro danou vlastnost) mohou přirozeně obsahovat více než jednu odlehlou hodnotu uvnitř jednotky zapříčiněnou instrumentálními vlivy. Po prozkoumání příčin odlehlých hodnot, může být vhodné přistoupit k vyloučení více než jednoho takového pozorování.

### 7.7.2 Inspekce trendů ve zpracování

Vynesení středních hodnot jednotek v pořadí v jakém byly zpracovávány, představuje jednoduchou metodu jak zkontrolovat vlivy programu zpracování. Alternativními metodami je vynesení reziduí oproti číslu jednotky po aplikaci modelu, který zahrnuje vlivy v sérii měření. To může být zejména platné u vyváženého hierarchického návrhu, kde střední hodnoty jednotek mohou být ovlivněny rozdíly v měřicí sérii. Použití matematických korekcí pro odstranění trendů způsobených zpracováním referenčního materiálu není vhodné. Pokud je zjištěn takový trend a daný materiál se stále pokládá za vhodný, musí se nejistota homogenity odhadnout takovým způsobem, aby zahrнула pozorovaný trend a rovněž i náhodnou variabilitu mezi jednotkami vyskytující se s tímto trendem.

### 7.7.3 Vyhodnocení členu mezi jednotkami – základní návrh

Základní návrh experimentů popsany v odstavci 7.6.2 se může analyzovat pomocí jednofaktorové analýzy rozptylu, jak je popsáno v příloze odstavce B1. Výpočet předpokládá, že existuje  $a$  jednotek RM a  $n_i$  měření jednotky RM  $i$ , kde  $i = 1 \dots a$ . Obvykle jednoduché studie homogenity představují vyvážené návrhy se stejným počtem  $n_0$  replikovaných měření u všech jednotek RM, takže  $n_1 = n_2 = \dots = n_a = n_0$ . Je-li pak použita jednofaktorová analýza rozptylu jak ji popisuje dodatek B1, směrodatnou odchylku mezi jednotkami  $S_{bb}$  lze vypočítat z průměrného čtverce mezi skupinami  $M_{\text{between}}$ , průměrného čtverce uvnitř skupin  $M_{\text{within}}$  a počtu replikátů u každé jednotky  $n_0$  podle (symboly jako v příloze B).

$$S_{bb}^2 = \max\left(\frac{M_{\text{between}} - M_{\text{within}}}{n_0}, 0\right) \quad (4)$$

V takovém případě rozptyl mezi jednotkami  $S_{bb}^2$  je identický s (druhou mocninou) příspěvku homogenity mezi jednotkami k nejistotě,  $u_{bb}^2$  (viz 7.11 [26]).

**POZNÁMKA 1** V příloze C, příkladu C1 se popisuje příklad testování homogenity za použití jednofaktorové analýzy rozptylu.

**POZNÁMKA 2** Často se dá předpokládat, že pravé střední hodnoty jednotek RM mají normální rozdělení, i když tento předpoklad není podmínkou výpočtu podle vzorce (4). Podobně lze často předpokládat, že náhodná chyba měření uvnitř jednotek má normální rozdělení.<sup>[13]</sup>

consistent with the remainder of the data set and that the number of units remaining in the study is sufficient.

**NOTE 3** Larger homogeneity studies (for example, 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 may be particularly useful after a balanced nested design as unit means may be influenced by measurement run differences. It is not appropriate to apply mathematical corrections to remove trends arising from the processing of the reference material. If such a trend is detected and the material is still found to be useful, the uncertainty of homogeneity has to 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 may be analysed by one-way analysis of variance as described in Clause B.1. The calculation assumes there are  $a$  RM units and  $n_i$  measurements on RM unit  $i$ , where  $i = 1 \dots a$ . 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  $n_1 = n_2 = \dots = n_a = n_0$ . If a one-way analysis of variance approach is then used as described in Clause B.1, the between-unit standard deviation  $S_{bb}$  can be computed from the between-group mean square  $M_{\text{between}}$ , the within-group mean square  $M_{\text{within}}$  and the number of replicates per unit  $n_0$  using (with symbols as in Annex B).

$$S_{bb}^2 = \max\left(\frac{M_{\text{between}} - M_{\text{within}}}{n_0}, 0\right) \quad (4)$$

In this simple case, the between-unit variance  $S_{bb}^2$  is identical to the (squared) between-unit homogeneity contribution to the uncertainty,  $u_{bb}^2$  (see 7.11 [26]).

**NOTE 1** Annex C, Example C.1 describes an example of a homogeneity test using one-way analysis of variance.

**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.<sup>[13]</sup>

### 7.7.4 Vyhodnocení členu mezi jednotkami – znáhodněný blokový návrh

Znáhodněný blokový návrh s  $a$  jednotkami RM s  $n_0$  replikáty, s jedním replikátem u jednotky v sérii a  $n_r$  sériemi, může být analyzován dvoufaktorovou analýzou rozptylu bez opakování. Statistický model je následující

$$x_{ij} = \mu + A_i + B_j + \varepsilon_{ij} \quad (5)$$

a je identický s modelem pro základní návrh (dodatek B.1) kromě přidání  $B_j$ , vlivu série  $j$  na výsledky. Stanovení rozptylu mezi jednotkami  $s_{bb}^2$  nezávisí na rozdělení  $A$  nebo  $B$ , i když velké vlivy v sérii by mohly v zásadě mít nepřímo vliv na reziduální člen, pokud preciznost závisí na pozorované úrovni střední hodnoty. Obvykle však vlivy v sérii, které jsou dostatečně veliké, aby měly takový vliv, aby činily takovou studii veskrze neplatnou. Kromě toho, analýza předpokládá, že vlivy v sérii a vlivy v jednotkách jsou nezávislé a aditivní. To je též rozumná aproximace relativně malých vlivů (ve srovnání se směrodatnou odchylkou uvnitř skupin a mezi skupinami) jak se očekává u testování homogenity.

Analýza rozptylu použitá u znáhodněného blokového návrhu, podle modelu ze vztahu (5) a s jedním pozorováním u každé jednotky a v každé sérii, vede k průměrnému čtverci mezi sériemi  $M_B$  spolu jako dříve, se středním čtvercem mezi jednotkami  $M_A$ , a reziduálním průměrným čtvercem  $M_{within}$ .

Tabulka analýzy rozptylu obsahuje též přidružené stupně volnosti pro každý člen. Reziduální průměrný čtverec  $M_{within}$  představuje nevychýlený odhad rozptylu opakovatelnosti  $s_r^2$ , i když v tabulce ANOVA se bude uvádět, že počet stupňů volnosti je roven  $(a-1)(n_0-1)$  spíše než  $a(n_0-1)$  jako u základního návrhu. Směrodatná odchylka mezi jednotkami  $s_{bb}$  se vypočítá podle vzorce (6) a má  $a-1$  stupňů volnosti.

$$S_{bb}^2 = S_A^2 = \max\left(\frac{M_A - M_{within}}{n_0}, 0\right) \quad (6)$$

POZNÁMKA Příklad zpracování znáhodněného blokového návrhu lze nalézt v příloze C, příklad C.3.

### 7.7.5 Vyhodnocení členu mezi jednotkami – vyvážený hierarchický návrh

Složky rozptylu u vyvážených hierarchických návrhů s  $a$  jednotkami rovnoměrně rozdělenými do  $n_r$  sérií a s  $m_w$  jednotkami v sérii se mohou vypočítat pomocí analýzy rozptylu popsané v ISO 5725-3,<sup>[14]</sup> která zahrnuje hierarchické návrhy experimentů nebo ve standardních statistických textech jako je odkaz <sup>[15]</sup>. Statistický model je následující

$$x_{ijk} = \mu + A_{ij} + B_j + \varepsilon_{ijk} \quad (7)$$

kde  $x_{ijk}$  je  $k$ 'tému pozorování jednotky  $i$  ( $i = 1..m_w$ ) v sérii  $j$ , ( $j = 1..n_r$ ).  $A_{ij}$  je vliv jednotky  $i$  v sérii  $j$ ,  $B_j$  je vliv v sérii  $j$  a  $\varepsilon_{ijk}$  ( $k = 1..n_0$ ) představuje obvyklý člen reziduální chyby. Analýza rozptylu u vyváženého

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

A randomized block design involving  $a$  RM units with  $n_0$  replicates, with one replicate per unit per run and  $n_r$  runs, may be analysed using two-way analysis of variance without replication. The statistical model is

$$x_{ij} = \mu + A_i + B_j + \varepsilon_{ij} \quad (5)$$

which is identical to the model for the basic design (Clause B.1) except for the addition of  $B_j$ , the effect of run  $j$  on the results. The determination of the between-unit variance  $s_{bb}^2$  is not dependent on the distribution of  $A$  or  $B$ , though large run effects could 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 to the within- and between-group standard deviation) expected in homogeneity testing.

Analysis of variance 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  $M_B$  together with, as before, a between-unit mean square  $M_A$ , and a residual mean square  $M_{within}$ .

The analysis of variance table also includes associated degrees of freedom for each term. The residual mean square  $M_{within}$  is an unbiased estimate of the repeatability variance  $s_r^2$ , 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 design. The between-unit standard deviation  $s_{bb}$  is calculated as in Formula (6) and has  $a-1$  degrees of freedom.

$$S_{bb}^2 = S_A^2 = \max\left(\frac{M_A - M_{within}}{n_0}, 0\right) \quad (6)$$

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

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

Variance components for balanced nested designs using  $a$  units equally distributed among  $n_r$  runs and with  $m_w$  units per run may be calculated using analysis of variance methods described in ISO 5725-3,<sup>[14]</sup> which includes nested designs, or by following standard statistical texts such as Reference <sup>[15]</sup>. The statistical model is

$$x_{ijk} = \mu + A_{ij} + B_j + \varepsilon_{ijk} \quad (7)$$

where  $x_{ijk}$  is the  $k$ 'th observation for unit  $i$  ( $i = 1..m_w$ ) in run  $j$ , ( $j = 1..n_r$ )  $A_{ij}$  is the effect of unit  $i$  in run  $j$ ,  $B_j$  is the effect of run  $j$  and  $\varepsilon_{ijk}$  ( $k = 1..n_0$ ) the usual residual error term. Analysis of variance for a balanced nested

hierarchického návrhu s jednotkami uspořádanými v sériích opět vytváří průměrný čtverec mezi sériemi  $M_B$  spolu s průměrným čtvercem mezi jednotkami  $M_A$ , a reziduální průměrný čtverec  $M_w$ . Za předpokladu jednotek uspořádaných v sériích a s  $n_0$  replikáty u každé jednotky člen mezi jednotkami se odhadne opět pomocí vzorce (4), i když s  $a - n_r$  stupni volnosti namísto  $a - 1$ , jak je tomu u základního návrhu.

Dvoufaktorová ANOVA pro hierarchický návrh je popsána v dodatku B.2 přílohy B. Pro analýzu studie homogeneity s  $a$  jednotkami rovnoměrně rozdělenými do  $n_r$  sérií a s  $m_w$  pozorováními u každé jednotky RM,  $a$ ,  $n_r$  a  $m_w$  nahrazují  $p$  všechny  $b_i$  a všechny  $n_{ij}$  z odstavce B.2.

### 7.7.6 Jiné návrhy pro homogenitu a alternativní metody odhadu

Jiné návrhy experimentů obvykle vyžadují pro analýzu použití specializovaného statistického softwaru schopného analýzy složek rozptylu pro širokou paletu návrhů. Některé z takových softwarů pokračují ve využití výpočtu průměrných čtverců pro odhad rozptylu a činí rovnými průměrné čtverce s očekávanými průměrnými čtverci; to je postup použitý v 7.7.3 až 7.7.5 výše. Dostupnými se staly další statistické metody, největšímu zájmu se těší metody maximální věrohodnosti a Bayesovská analýza. Ačkoliv podrobné výpočty jsou mimo rámec tohoto Pokynu, uvádíme níže určitý návod pro výběr takových metod.

Metody maximální věrohodnosti patří mezi nejobecnější a široce používané metody odhadu složek rozptylu. Doporučovaný postup pro odhad rozptylu při práci s referenčními materiály je variantou známou jako restringovaný (nebo někdy 'reziduální') maximálně věrohodnostní odhad (REML). Ten poskytuje v podstatě identické výsledky jako analýza rozptylu u vyvážených návrhů, kde nechybí žádná data, ale též dobře zpracovává komplexní návrhy, chybějící data a znatelnou nevyváženost. REML se může tudíž použít na místo metod založených na ANOVA uvedených výše a rovněž i v širším rozsahu okolností.

Bayesovská analýza nabírá na popularitě s tím, jak se staly dostupnými pokročilé výpočetní metody. Bayesovská analýza však závisí na použití apriorního rozdělení pravděpodobnosti odhadovaných parametrů. Též obecně vyžaduje pečlivou specifikaci statistického modelu a obecně je ve zvýšené míře výpočtově náročná. Ačkoliv Bayesovské metody poskytují ve zkušených rukou dobré výsledky, mohou též poskytovat neočekávané a zavádějící výsledky pokud je nesprávně specifikován nebo jsou zvolena nevhodná výchozí nastavení zejména u poměrně malých souborů dat, které jsou běžné u certifikace referenčních materiálů. Bayesovské metody pro odhad složek rozptylu by měly být tudíž používány za pomoci kvalifikovaného statistika se zkušenostmi z Bayesovské analýzy malých datových souborů.

design with units nested in runs will again generate a between-run mean square  $M_B$  together with a between-unit mean square  $M_A$ , and a residual mean square  $M_w$ . Assuming units nested within runs and with  $n_0$  replicates per 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 Clause B.2. For analysing a homogeneity study using  $a$  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 Clause B.2.

### 7.7.6 Other homogeneity designs and alternative estimation methods

Other designs will normally 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 above. Other statistical methods have become available; however, of most interest are maximum likelihood methods and Bayesian analysis. Although detailed calculations are beyond the scope of this Guide, some guidance is given below for the selection of such methods.

Maximum likelihood methods are among the most general and widely used methods of variance component estimation. The recommended approach for estimating variances in reference material work is the variant known as Restricted (or, sometimes, 'residual') Maximum Likelihood estimation (REML). This provides essentially identical results to those of analysis of variance 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 above 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 reference material certification. Bayesian methods for variance component estimation in reference material certification should therefore be used only with the assistance of a qualified statistician experienced in Bayesian analysis of small data sets.

## 7.8 Nedostatečná opakovatelnost měřicí metody

Vždy není proveditelné uskutečnit studii homogenity s použitím měřicí metody, která má dostačující opakovatelnost (tedy vyhovuje požadavku 7.5). V těchto případech se může použít alternativní přístup, kde se odhadne horní mez členu mezi jednotkami.

Pojednání o různých přístupech jak získat odhad nejistoty v případě nedostatečné opakovatelnosti měřicího postupu obsahuje odkaz <sup>[16]</sup>. Příkladem aplikace doporučení z odkazu <sup>[16]</sup> je příklad C.2 přílohy C.

## 7.9 Homogenita uvnitř jednotek

### 7.9.1 Posouzení potřeby studie homogenity uvnitř jednotek

Homogenita uvnitř jednotek je problémem, který vzniká, když minimální odběr vzorku specifikovaný v návodu pro použití je menší než velikost kompletní jednotky RM. Je vždy důležité zajistit, že specifikovaný odběr vzorku je vyhovující a homogenita mezi jednotkami sama o sobě nemůže poskytovat takovou záruku. Tak kde minimální odběr vzorku je malý a existuje vysoké riziko znatelné heterogenity uvnitř jednotky, tam by se měla provést experimentální studie homogenity uvnitř jednotek. Situace, které znamenají vysoké riziko znatelné heterogenity uvnitř jednotek, jsou jedna nebo více z následujících:

- minimální odběr vzorků je podstatně menší než velikost jednotky,
- materiál, který se připravuje smísením prášků nebo zrnitých materiálů,
- materiál, který se připravuje vmísením malého množství jedné komponenty do výchozí matrice, nebo
- materiál s předtím známou heterogenitou uvnitř jednotek.

Experimentální hodnocení může spočívat v testování na významnou heterogenitu uvnitř jednotek nebo určení minimálního odběru vzorku.

### 7.9.2 Testování významné heterogenity uvnitř jednotky

Testování významné heterogenity uvnitř jednotek používá návrh experimentů podobný jako testování homogenity mezi jednotkami, s tím rozdílem, že sledovaný rozptyl představuje člen uvnitř jednotky, mezi podvzorky. Na obrázku 5 jsou znázorněny některé možné návrhy experimentů pro provedení studie homogenity uvnitř jednotek. V nejjednodušším z nich [obr. 5 a)], se odebere z jedné jednotky RM  $m$  zkoušených vzorků o velikosti v rozsahu zamýšlené hodnoty minimálního odběru vzorku a každý je zpracován a změřen  $n_w \geq 1$  krát. Tam, kde není možno odebrat potřebné množství alikvotů z jedné

## 7.8 Insufficient repeatability of the measurement method

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

A discussion about various approaches to obtain an uncertainty estimate that accounts for insufficient repeatability of the measurement procedure is given in Reference <sup>[16]</sup>. An example of application of the recommendations of Reference <sup>[16]</sup> is given as Annex C, Example 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 intake 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 sample intake is sufficient, and a between-unit homogeneity study alone does not provide such assurance. Where the minimum sample intake 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 intake substantially less 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; or
- a material with previously known within-unit heterogeneity.

Experimental assessment may consist of a test for significant within-unit heterogeneity or a determination of minimum sample intake.

### 7.9.2 Testing for significant within-unit heterogeneity

A test for significant within-unit heterogeneity uses a design similar to between-unit homogeneity testing, with the exception that the variance of interest is the within-unit, between-subsample term. In Figure 5 shows some possible designs for a within-unit homogeneity study. In the simplest [in Figure 5 a)],  $m$  test portions, in the size range of the envisaged minimum sample intake, 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 in order to obtain more representative information, more RM units can be

jednotky RM, nebo pro získání reprezentativnější informace, může se zahrnout více jednotek RM, typicky se sníženou replikací uvnitř jednotek [obr. 5 b)].

Testování heterogenity uvnitř jednotky by mělo vykazovat alespoň pět stupňů volnosti pro člen uvnitř jednotky. Toho lze dosáhnout například odběrem šesti nebo více podvzorků z jediné jednotky [ $m \geq 6$  na obr. 5 a)], odběrem tří podvzorků z každé ze tří jednotek [ $a = 3, m = 3$  na obr. 5 b)]; nebo odběrem duplikátních podvzorků z každé z pěti nebo více jednotek [ $a \geq 5, m = 2$  na obr. 5 b)]. Tam, kde se odebírají podvzorky z několika jednotek, analýza by měla zahrnovat i člen mezi jednotkami způsobem, který je analogický zahrnutí členu odpovídajícího sérii u vyváženého hierarchického návrhu v 7.7.5.

Tam kde se provádí opakované analýzy u každého podvzorku, měla by se používat analýza rozptylu s následným F testem pro významný rozptyl mezi podvzorky. U jednotlivé RM jednotky s alikvoty měřeními několikanásobně se může použít jednofaktorová ANOVA s podvzorkem jako skupinovým faktorem jako v dodatku B.1. Tam kde se podvzorky odebírají jako násobné alikvoty z několika jednotek jako na obr. 5 b) dvoufaktorová ANOVA pro hierarchické návrhy je vhodným řešením (dodatek B.2). Pokud se může uskutečnit pouze jedno měření u každého podvzorku  $n_w = 1$  rozptyl pozorování (nebo z několika jednotek, člen uvnitř skupiny u jednofaktorové analýzy rozptylu) by se měl porovnat se spolehlivým odhadem opakovatelnosti metody, obvykle za použití F testu na významně vyšší rozptyl výsledků z různých podvzorků. V případě kdy opakovatelnost je dobře stanovena s vysokými stupni volnosti je takové hodnocení možno provést porovnáním pozorovaného rozptylu se známým rozptylem opakovatelnosti použitím  $X^2$ -testu.

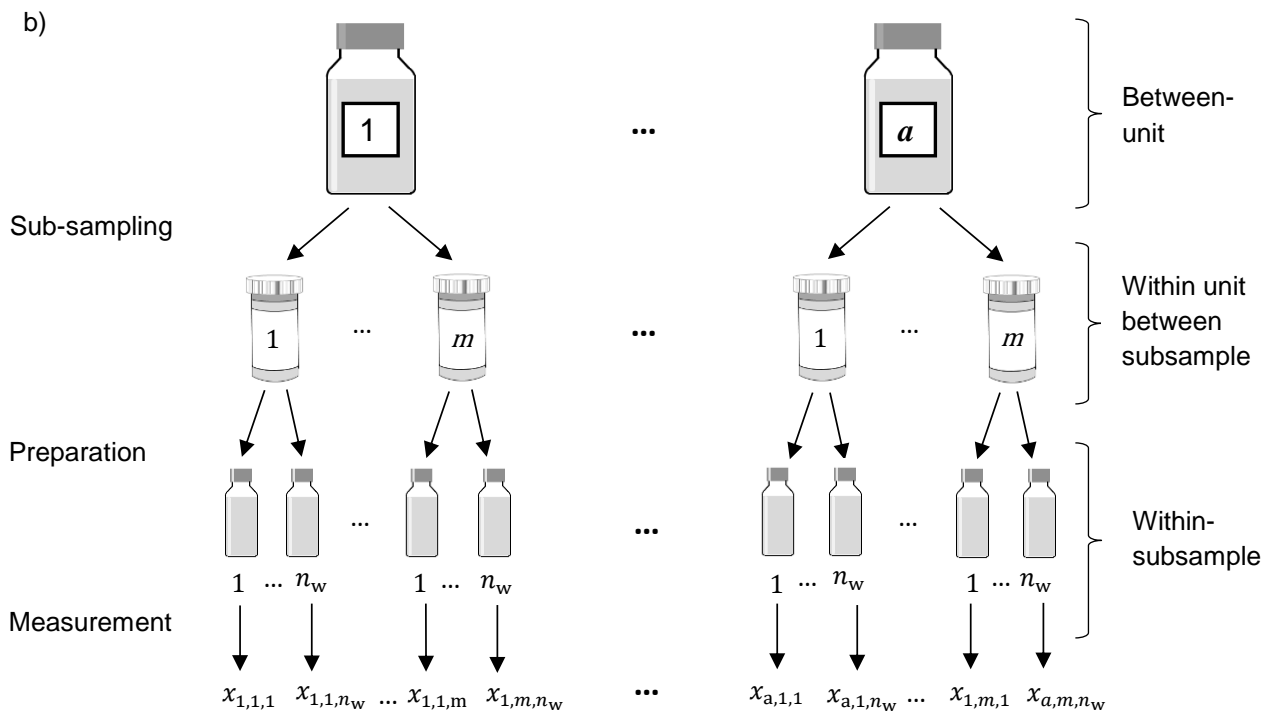
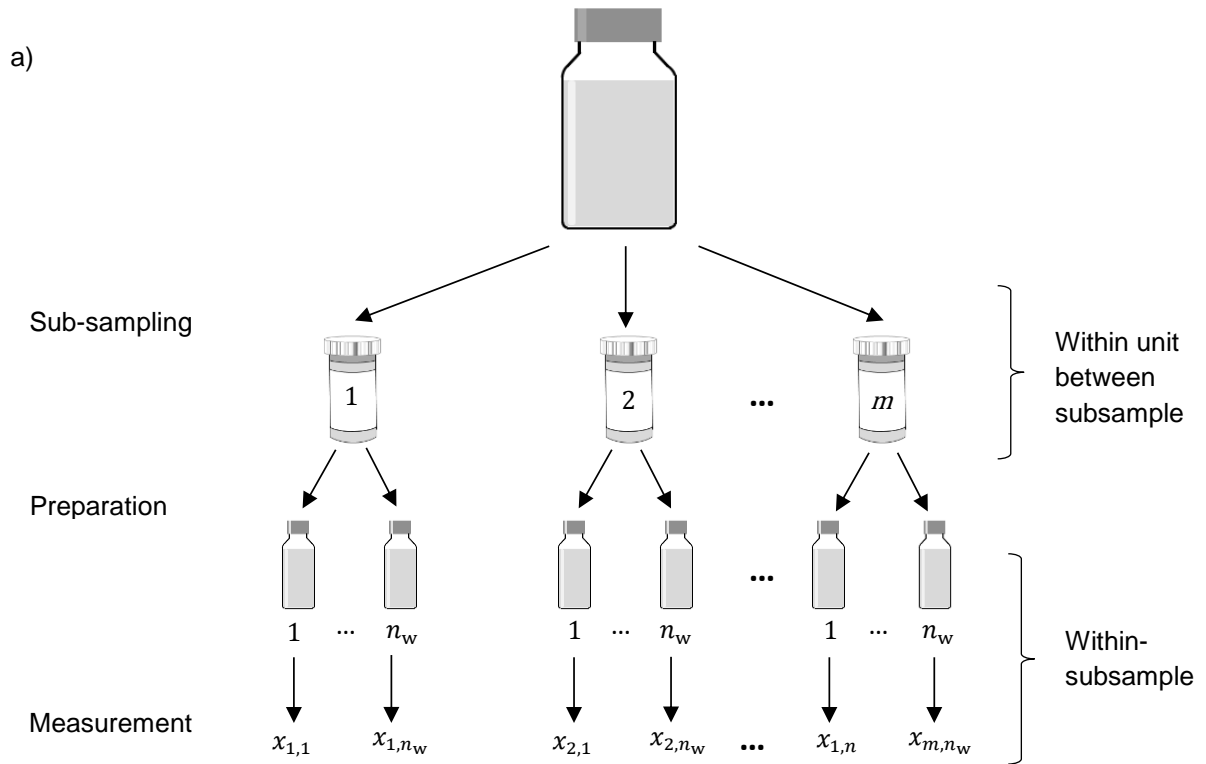
**POZNÁMKA** I tam kde je možno uskutečnit opakovaná měření podvzorku, není vždy možné provést opakovanou přípravu podvzorku jak se uvádí na obr. 5 b), zvláště pokud velikost podvzorku byla zvolena na úrovni navrhovaného minimálního odběru vzorku. To pak má za následek návrh ležící mezi těmi, které znázorňují obrázky 5a) a 5 b).

included, typically with reduced within-unit replication [Figure 5 b)].

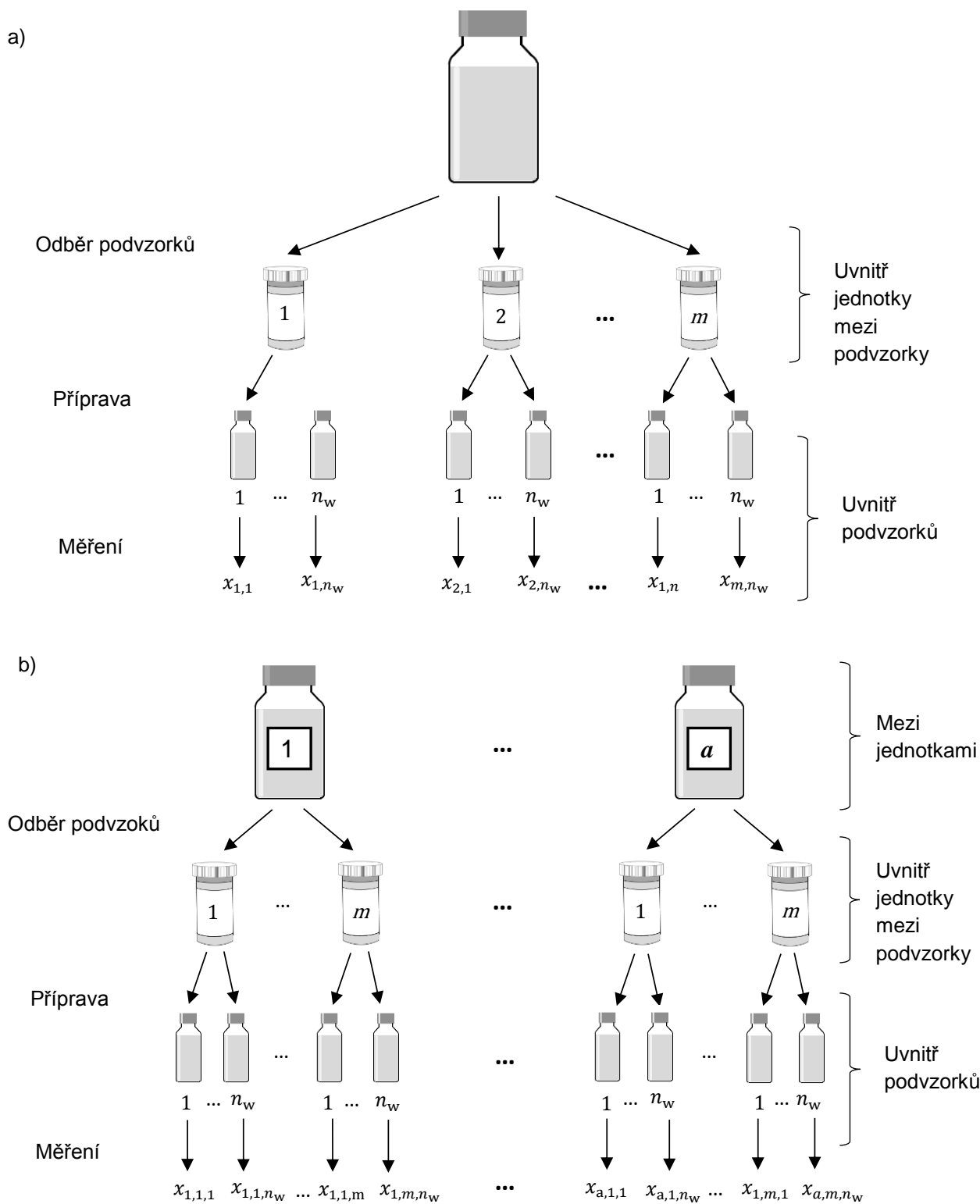
A test for within-unit heterogeneity should provide at least five degrees of freedom for the within-unit term. This can be achieved by, for example, taking six or more subsamples from a single unit [ $m \geq 6$  in Figure 5 a)], taking three subsamples from each of three units [ $a = 3, m = 3$  in Figure 5 b)]; or by taking duplicate subsamples from each of five or more units [ $a \geq 5, m = 2$  in Figure 5 b)]. Where subsamples are taken from multiple 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, analysis of variance followed by an F-test for significant between-subsample variance may 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 Clause B.1; where multiple units are subsampled and multiple aliquots taken, as in Figure 5 b), two way ANOVA for nested designs is appropriate (Clause B.2). Where only a single observation can be made per subsample  $n_w = 1$ , the variance of the observations (or, from multiple units, the within-group term from one-way analysis of variance) should be compared with a reliable estimate of the method repeatability, usually using an F-test for significantly greater variance for results on different subsamples. For the case 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  $X^2$ -test.

**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 intake. This results in a design intermediate between those shown in Figure 5 a) and in Figure 5 b).





**Klíč**

- a ...Jediná jednotka RM, replikovaná příprava a měření  
 b ...Několik jednotek RM replikovaná příprava a měření

**Key**

- a Single RM unit, replicated preparation and measurement  
 b Multiple RM units, replicated preparation and measurement

**Figure 5 — Schematic designs for within-unit homogeneity study****Obrázek 5 — Schématické návrhy experimentů pro studii homogenity uvnitř jednotek**

## 7.9.3 Hodnocení minimálního odběru vzorku

### 7.9.3.1 Experimentální stanovení

Minimální odběr vzorku lze stanovit experimentálně provedením experimentální studie homogenity uvnitř jednotky s různými velikostmi odebraného vzorku. Jelikož směrodatná odchylka homogenity uvnitř jednotky závisí na počtu částic, které jsou nositeli určité vlastnosti, očekává se, že pozorovaná směrodatná odchylka se bude zvyšovat se snižujícím se odebraným množstvím a bude se blížit opakovatelnosti zkušební metody se zvyšujícím se odběrem vzorku. Toto může tvořit základ studie odběru vzorku, například v následujících krocích:

- stanovte pozorovanou směrodatnou odchylku pro replikáty v rozsahu odlišných odběrů vzorku,
- vyneste pozorovanou směrodatnou odchylku oproti odběru vzorku,
- buď
  - použijte graf k prokázání, že velikost odběru vzorku nemá vliv na pozorovanou směrodatnou odchylku v hodnoceném rozsahu, nebo
  - zvolte odběr vzorku, při kterém odhadnutá směrodatná odchylka se neliší od směrodatné odchylky opakovatelnosti zkušební metody.

Protože počet stupňů volnosti je malý, je důležité, aby rozsah studovaných odběrů vzorku vykázal významný rozdíl, pokud bude existovat znatelná heterogenita uvnitř jednotky. V takovém experimentu odběry vzorků v rozsahu nejméně pětinasobku (například od 10 mg do 50 mg) by se měly použít, tam, kde je to možné.

Tam, kde lze prokázat, že rozptyl mezi podvzorky je nepřímo úměrný odebranému množství, použije se Ingamellova teorie vzorkování.<sup>[17]</sup> Za těchto podmínek je hmotnost  $m_1$  požadovaná k dosažení konkrétní relativní směrodatné odchylky  $s'$  dána jako

$$m_1 = K_1/s'^2 \quad (8)$$

kde  $K_1$  je Ingamelova vzorkovací konstanta. [Ingamelova vzorkovací konstanta je hmotnost materiálu, která vede k relativní směrodatné odchylce mezi podvzorky rovné 0,01 (1 %) a může se stanovit z relativní směrodatné odchylky mezi podvzorky při jiných hmotnostech.]

### 7.9.3.2 Jiné metody

Minimální odběr vzorku se může nastavit na základě jiných dat nebo podle zkušenosti. Mezi jiná data, která mohou sloužit pro volbu minimálního odběru vzorku, patří například:

- odběry vzorku (navážky) používané k získání výsledků s přijatelnou precizností ve studii stability,

## 7.9.3 Assessing minimum sample intake

### 7.9.3.1 Experimental determination

The minimum sample intake may be determined experimentally by carrying out a within-unit homogeneity study for different sizes of sample intake. As the within-unit homogeneity standard deviation depends on the number of particles carrying a certain property, the observed standard deviation is expected to increase as the sample intake decreases and to converge to the test method repeatability as the sample intake increases. This can form the basis for a sample intake study, for example, by following these steps:

- determine the observed standard deviation of replicates at a range of different sample intakes;
- plot the observed standard deviation against the sample intake;
- either
  - use the plot to demonstrate that the sample intake has no effect on the observed standard deviation in the range assessed, or
  - choose a sample intake at which the estimated standard deviation is indistinguishable from the repeatability standard deviation for the test method.

Since the number of degrees of freedom is small, it is important that the range of sample intakes studied be sufficient to yield a significant difference if there is appreciable within-unit heterogeneity. Sample intakes covering a range of at least a factor of five (for example, from 10 mg to 50 mg) should therefore be included in such an experiment, where possible.

Where it can be shown that the between-subsample variance is inversely proportional to the mass taken, Ingamell's sampling theory applies.<sup>[17]</sup> In this circumstance, the mass  $m_1$  required to achieve a particular between-subsample relative standard deviation  $s'$  is given by

$$m_1 = K_1/s'^2 \quad (8)$$

where  $K_1$  is Ingamell's sampling constant. [Ingamell's sampling constant is the mass of material that leads to a between-subsample relative standard deviation of 0,01 (1 %) and can be determined from the between-subsample relative standard deviation at other subsample masses.]

### 7.9.3.2 Other methods

Minimum sample intake may be set based on other data or by experience. Other data that may inform the choice of minimum sample intake includes, for example:

- sample intake used for obtaining results with acceptable precision in a stability study;

- odběry vzorku poskytující přijatelnou preciznost při charakterizační studii prováděné v jediné laboratoři,
  - odběry vzorků z laboratoří dosahujících přijatelnou preciznost při mezilaboratorní charakterizaci,
  - odběry vzorků používané při studii homogenity.
- sample intake providing acceptable precision in a characterization study carried out in a single laboratory;
  - sample intake from laboratories achieving acceptable precision in an interlaboratory characterization;
  - sample intake used for the homogeneity study.

### 7.10 Kontrola dostatečné homogenity

Po stanovení směrodatné odchylky mezi jednotkami a (tam, kde je to vhodné) uvnitř jednotek měl by výrobce potvrdit, zda variabilita uvnitř jednotek a mezi nimi je dostatečně malá pro zamýšlené použití daného materiálu. Takové potvrzení může představovat, například:

- a) porovnání směrodatné odchylky mezi jednotkami a/nebo uvnitř jednotek s nejistotou charakterizace, aby se potvrdilo, že směrodatná odchylka (odchylky) jsou malé ve srovnání s nejistotou charakterizace.
- b) výpočtem kombinované nejistoty certifikované hodnoty se zohledněním heterogenity a potvrzením, že získaná nejistota certifikované hodnoty je přijatelná pro zamýšlené použití,
- c) u referenčních materiálů, u kterých se nebude počítat kombinovaná nejistota, zkontrolováním že směrodatná odchylka mezi jednotkami  $s_{bb}$  je malá ve srovnání s typickou mezilaboratorní směrodatnou odchylkou reprodukovatelnosti  $s_r$  v oblasti použití. V ideálním případě  $s_{bb}$  by měla být menší než  $s_r/3$ ,
- d) potvrzení například použitím F testu, že člen mezi jednotkami není statisticky významný na 95 % konfidenční úrovni.

**POZNÁMKA** Postrádání statistické významnosti neprokazuje neexistenci heterogenity, ale může představovat užitečné praktické kritérium v případech, kde není záměrem poskytovat nejistotu k hodnotě vlastnosti a metoda pro testování homogenity je jednou z typických metod používaných koncovými uživateli materiálu.

### 7.11 Vyhodnocení nejistoty ze studie homogenity

Nejistota certifikované hodnoty (viz 10.2) by měla zahrnovat příspěvek  $u_{hom}$  za jakoukoliv detekovanou heterogenitu. Příspěvek  $u_{hom}$  by neměl být menší než směrodatná odchylka mezi jednotkami  $s_{bb}$  stanovená pomocí experimentálních studií homogenity.  $s_{bb}$  se může rovnat nule nebo být zanedbatelná ve srovnání s dalšími nejistotami, v jejichž případě může být člen  $u_{hom}$  vynechán ledaže by heterogenita uvnitř jednotek byla významná.

Tam, kde existuje významná heterogenita uvnitř jednotek při minimálním odběru vzorku doporučeném v návodu k použití, nejistota přidružená k certifikované hodnotě by se měla zvýšit začleněním dalšího příspěvku  $u_{wb}$ , který by neměl být

### 7.10 Check for sufficient homogeneity

Having determined the between-unit and (where appropriate) within-unit standard deviation, the producer should confirm that the variation within and between units is sufficiently small for the intended use of the material. Such confirmation may include, for example:

- a) comparison of the between-unit and/or within-unit standard deviation to the uncertainty associated with characterization to confirm that the standard deviation(s) are small compared to the characterization uncertainty;
- b) calculation of the combined uncertainty of the certified value, with allowance for heterogeneity, and confirmation that the uncertainty of the certified value is acceptable for the intended use;
- c) for reference materials where the combined uncertainty will not be calculated, checking that the between-unit standard deviation  $s_{bb}$  is small compared to the typical interlaboratory reproducibility standard deviation  $s_r$  in the field of use. Ideally  $s_{bb}$  should be less than  $s_r/3$ ;
- d) confirmation, for example by use of an F test, that the between-unit term is not statistically significant at the 95 % level of confidence.

**NOTE** Lack of statistical significance does not prove the absence of heterogeneity, but can be a useful practical criterion when it is not intended to provide an uncertainty with a property value and the method used for the homogeneity test is typical of methods employed by end users of the material.

### 7.11 Uncertainty evaluation from homogeneity studies

The uncertainty for a certified value (see 10.2) should include an allowance  $u_{hom}$  for any detected heterogeneity.  $u_{hom}$  should not be less than the between-unit standard deviation  $s_{bb}$  determined from experimental homogeneity studies.  $s_{bb}$  can be zero or negligible compared to other uncertainties, in which case the term  $u_{hom}$  may be omitted unless within-unit heterogeneity is important.

Where there is significant within-unit heterogeneity at the minimum sample intake to be recommended in the instructions for use, the uncertainty associated with a certified value should be increased by the inclusion of a further allowance  $u_{wb}$  that should not be less than

nižší než člen mezi podvzorky odvozený z experimentálních studií (viz 7.9). Nejistota  $u_{\text{hom}}$  přidružená heterogenitě je pak dána jako

$$u_{\text{hom}} = \sqrt{u_{\text{bb}}^2 + u_{\text{wb}}^2} \quad (9)$$

Příspěvek  $u_{\text{hom}}$  k nejistotě certifikované hodnoty by se mohl též zvýšit, aby zahrnul nedostatečnou preciznost zkušební metody použité pro testování homogenity, jak je uvedeno v 7.8.

## 8 Hodnocení a monitorování stability

### 8.1 Úvodní část

Stabilita předmětných vlastností je jedním z klíčových rysů všech referenčních materiálů. Je důležité, aby hodnota pro každou předmětnou vlastnost byla v době použití konzistentní s hodnotou udávanou na certifikátu nebo jiné dokumentaci doprovázející materiál.

Hodnota každé vlastnosti se může měnit s časem z mnoha důvodů, v různém stupni a různou rychlostí v závislosti na podmínkách. Tři druhy podmínek jsou zvláště důležité: podmínky dlouhodobého skladování v zařízeních výrobce, podmínky během přepravy do prostor uživatelů a specifikované podmínky skladování a používání v prostorách uživatelů.

Způsob a rychlost změny se může značně lišit u různých materiálů. Některé se mění málo nebo vůbec ne v širokém rozsahu podmínek. Některé se mění rychle za okolních podmínek a vyžadují skladování při nízkých teplotách. Forma degradace se může lišit: některé materiály se mění téměř lineárně po dlouhou dobu, některé mohou podléhat autokatalytickým nebo méně předvídatelným rychlým změnám po období stability. Některé se mohou rychle měnit v průběhu počáteční doby po zpracování a pak zůstat stabilní po dlouhou dobu. U některých se změna může týkat některých jednotek materiálu více než jiných, což vede ke změnám homogenity mezi jednotkami spíše než ke konzistentní změně u všech jednotek.

Tyto velmi odlišné způsoby změn, za různých podmínek se mohou těžko předpovídat i po rozsáhlé experimentální studii. Existuje tudíž vždy určité riziko, že hodnota jedné nebo více vlastností se bude neočekávaně měnit během životnosti referenčního materiálu, nebo i během životnosti konkrétní šarže referenčního materiálu, který se pravidelně vyrábí. Proto se očekává, že výrobce bude řídit zpracování, skladování, balení, podmínky přepravy, monitorování po certifikaci a poradní servis koncovým uživatelům takovým způsobem, aby riziko neočekávané změny bylo tak malé jak je jen možné.

Výrobce by měl:

- hodnotit, experimentálně, jestliže je to nutné, stabilitu všech vlastností referenčního materiálu za plánovaných podmínek dlouhodobého skladování a

the between-subsample term derived from experimental studies (see 7.9). The uncertainty  $u_{\text{hom}}$  associated with heterogeneity is then given by

$$u_{\text{hom}} = \sqrt{u_{\text{bb}}^2 + u_{\text{wb}}^2} \quad (9)$$

The contribution  $u_{\text{hom}}$  to the uncertainty of a certified value may also be increased to allow for insufficient precision of the test method used for homogeneity testing, as provided for in 7.8.

## 8 Assessment and monitoring of stability

### 8.1 Preamble

Stability of the properties of interest is one of the key features of all RMs. It is important that the value for each property of interest is, at the time of use, consistent with the stated value on the certificate or other documentation accompanying the material.

The value of each property may change over time for a variety of reasons, to different degrees, and at different rates depending on the conditions. Three sets of conditions are particularly important: conditions during long-term storage at the producer's facilities, conditions during transport to the user's premises, and the specified conditions of storage and use at the user's premises.

The form and rate of change may differ considerably for different materials. Some change little or not at all under a wide range of conditions. Some change rapidly under ambient conditions and require low temperature storage. The form of degradation may differ: some materials change almost linearly over long periods; some may undergo autocatalytic or less predictable rapid change after a period of stability. Some may change rapidly during an initial period after processing and then remain stable over long periods. For some, change may affect some units of the material much more than others, leading to changes in between-unit homogeneity rather than a consistent change in value for all units.

These very different patterns of change, under different conditions, can be hard to predict even after extended experimental study. There is therefore, always some risk that the value of one or more properties will change unexpectedly during the life of a reference material, or even during the life of a particular batch of a reference material that is produced regularly. Accordingly, the producer is expected to manage material processing, storage, packaging, transport conditions, post-certification monitoring and advice to end users so that the risk of unexpected change is as small as reasonably possible.

The producer should:

- assess, by experimentation if necessary, the stability of all properties of a reference material under planned long-term storage conditions, if any,

- případně volit předběžnou úpravu, balení a skladovací podmínky podle toho,
- b) hodnotit, experimentálně jestliže je to nutné, stabilitu všech vlastností referenčního materiálu za plánovaných podmínek přepravy a volit takové přepravní podmínky, které vykazují udržet nízké riziko významných změn během přepravy,
- c) poskytovat pokyny týkající se skladování a používání daného materiálu, které jsou nutné pro udržení stability v prostorách uživatele,
- d) vytvořit plán monitorování stability materiálů skladovaných po dlouhou dobu, který umožní okamžitou detekci změny, berouce v úvahu možné rychlosti změny,
- e) tam, kde stabilita certifikované hodnoty nemůže být zajištěna, zohlednit v uváděné nejistotě možnost změny hodnoty před použitím a tam, kde změna v čase může být předpovězena, poskytnout způsob na korekci certifikované hodnoty a její nejistoty pro očekávanou změnu v čase,
- f) tam, kde je dovolen opakovaný odběr vzorků z jednotky referenčního materiálu podle návodu k použití, zhodnotit možné vlivy na stabilitu materiálu a přijmout vhodná opatření.
- and choose pre-treatment, packaging and storage conditions accordingly;
- b) assess, by experimentation, if necessary, the stability of all properties of a reference material under planned conditions of transport, and choose transport conditions that have been shown to maintain a low risk of significant change during transport;
- c) provide any advice on storage and use of the material that is necessary to maintain stability at the user's premises;
- d) select a scheme for monitoring the stability of materials held in long term storage that permits prompt detection of change, taking into account the possible rate of change;
- e) where the stability of a certified value cannot be ensured, make due allowance in the stated uncertainty for possible change in the value prior to use or where the change with time can be predicted, provide a means of correcting the certified value and its uncertainty for expected change over time;
- f) where repeated sampling from a reference material unit is permitted by the instructions for use, assess the possible effects on the stability of the material and take appropriate action.

V následujících odstavcích se poskytuje poučení o potřebě experimentální studie a v obvyklém případě, kdy budou experimentální studie potřeba, se poskytuje návod o jejich podstatě a provádění. Jsou poskytnuty též rady o nezbytnosti nejistot přidružených k možné nestabilitě a jejich hodnocení pro ty případy, kde zůstává znatelné riziko trvalé změny v čase.

The following subclauses provide guidance on the need for experimental study and, in the common case where experimental studies are required, provide guidance on their nature and conduct. Guidance is also provided on the necessity for, and assessment of, uncertainties associated with possible instability for those cases where there remains an appreciable risk of steady change over time.

**POZNÁMKA 1** Hodnocení stability podle a) a b) výše se vyžaduje ve shodě s Pokynem ISO 34.[2] Pokyn ISO 34 dále vyžaduje monitorování stability v případě potřeby a tam, kde degradace je detekovatelná, je též vyžadováno dodržování položek d) a e) uvedených výše.

**NOTE 1** Assessment of stability as at a) and b) above is required for compliance with ISO Guide 34.[2] ISO Guide 34 additionally requires stability monitoring where appropriate, and, where degradation is detectable, adherence to items d) and e) above is also required.

**POZNÁMKA 2** Očekává se, že pro většinu materiálů výrobci budou volit takovou kombinaci předběžné úpravy, balení, skladování, podmínek přepravy a plánu monitoringu, které povedou k tomu, že se bude přiměřeným způsobem očekávat jen zanedbatelná změna v čase.

**NOTE 2** It is expected that for most materials, producers will select a combination of pre-treatment, packaging, storage, transport conditions and a monitoring scheme that lead to a reasonable expectation of negligible change over time.

**POZNÁMKA 3** Výsledky monitoringu mohou dodatečně sloužit pro řízení stability stejných nebo podobných referenčních materiálů v budoucnosti.

**NOTE 3** The results of monitoring can additionally inform future stability management of the same, or related, RMs.

**POZNÁMKA 4** Jakýkoliv způsob korekce přidělených hodnot vlivem předpovídaných změn v čase budou u certifikovaných hodnot dodatečně vyžadovat odhad nejistoty přidružené k takové korigované hodnotě.

**NOTE 4** Any means of correcting assigned values for predicted change over time will, for certified values, additionally require the estimation of the uncertainty associated with the corrected value.

**POZNÁMKA 5** Materiál může být stabilní vůči jedné vlastnosti, i když se hodnoty dalších vlastností mohou při měření v průběhu času významně měnit.

**NOTE 5** A material can be stable with respect to one property, even though other property values can change appreciably when measured over time.

**POZNÁMKA 6** Většina statistických postupů pro data popisující změnu v čase používají matematický model pro odhad plynulé změny. Avšak i když je možno kontrolovat, zda model odpovídá pro období dané studie, velmi široký rozsah forem budoucí možné degradace zabraňuje spolehlivé předpovídání. Není ani praktické zahrnovat složky nejistoty zohledňující možné alternativní modely degradace. Velké rozdíly v chování různých modelů by činily tokové nejistoty neprakticky vysokými. Jediným spolehlivým způsobem v praxi, jak odhadovat změny v delších obdobích za daných podmínek je sledovat materiál po celou předemtnou dobu. Z tohoto důvodu, ať už je ve studii použit jakýkoliv model testování změny, zůstává monitorování stability nezbytnou záležitostí pro jakýkoliv materiál, určený k používání po delší období.

## 8.2 Hodnocení stability

### 8.2.1 Požadavek na hodnocení stability

U všech referenčních materiálů by měla být hodnocena stabilita. Hodnocení stability může spočívat v experimentální studii k odhadu zbývajícího stupně nestability kandidátského RM po zpracování nebo pro potvrzení stability materiálu. Každá detekovatelná změna v hodnotách vlastností vlivem dlouhodobé nestability nebo vlivu přepravy by se měla zvažovat při uvádění hodnot vlastností a odhady změn by se *měly* zohlednit (a ve shodě s Pokynem ISO 34,<sup>[2]</sup> *musí* zohlednit) v nejistotě certifikovaných hodnot.

Když se referenční materiál vyrábí v opakovaných šaržích, výrobce by měl dodatečně hodnotit riziko změn stability od jedné šarže ke druhé.

### 8.2.2 Typy (ne)stability

Výroby referenčních materiálů se týkají především dva typy (ne)stability:

- dlouhodobá stabilita daného materiálu (tj. stabilita materiálu v průběhu doby platnosti za specifikovaných podmínek skladování),
- stabilita za přiměřeně očekávaných podmínek přepravy („přepravní stabilita“).

Dlouhodobá stabilita referenčního materiálu se váže na chování RM na regálech výrobce, nebo když je skladován v souladu s předepsanými podmínkami u uživatele. Dlouhodobou stabilitu je třeba hodnotit před distribucí materiálu uživatelům.

Přepravní stabilita souvisí s dalšími mimořádnými vlivy při přepravě vzorků. V některých případech není možné udržet během přepravy vhodné podmínky se zřetelem na stabilitu daného RM a v takovém případě by se měl zohlednit příspěvek určité mimořádné nejistoty u certifikovaných hodnot. Výrobce by měl před distribucí materiálu uživatelům, zhodnotit stabilitu každé předemtné vlastnosti za přiměřeně očekávaných podmínek přepravy.

**NOTE 6** Most statistical treatments for data describing change over time use a mathematical model to estimate continuous change. However, while a model may be checked for adequacy across the time course of a given study, the very wide range of forms for possible future degradation prevent reliable prediction. Nor is it practical to include uncertainty components accounting for possible alternative degradation models; the large divergence in behaviour of different models would render such uncertainties impractically large. In practice, the only reliable way of estimating change over extended periods at a given set of conditions is to observe the material over the complete period of interest. For this reason, whatever model is used to test for change during the study and to predict future change, stability monitoring remains essential for any material intended to remain available for extended periods.

## 8.2 Assessment of stability

### 8.2.1 Requirement for stability assessment

The stability of all RMs should be assessed. The assessment of stability may involve an experimental study to estimate the remaining degree of instability of the candidate RM after processing or to confirm the stability of the material. Any detectable changes in the property values due to long-term instability or the effect of transportation should be considered in the claims for the property values, and the uncertainty for these estimates of change *should* (and, for compliance with ISO Guide 34,<sup>[2]</sup> *shall*) be included in the uncertainty for certified values.

Where a reference material is produced in repeated batches, the producer should additionally assess the risk of changes in stability from one batch to the next.

### 8.2.2 Types of (in)stability

Two types of (in)stability are particularly relevant in the production of reference materials:

- the long-term stability of the material (i.e. the stability of the material during the period of validity under specified storage conditions);
- the stability under reasonably expected conditions of transport (“transportation stability”).

The long-term stability of a reference material is associated with the behaviour of the RM on the shelves of the producer or when stored according to prescribed conditions by the user. Long-term stability of the material should be assessed prior to distribution of the material to users.

The transportation stability is associated with any *extra* effects due to transport of the samples. In some cases, it is not possible to maintain appropriate conditions with respect to the stability of the RM during transport and, in this case, allowance should be made for some extra uncertainty in certified values. The producer should, prior to distribution of the material to users, assess the stability of each property of interest under reasonably expected conditions of transport.

POZNÁMKA 1 Přepravní stabilita se často uvádí jako „krátkodobá stabilita“.

POZNÁMKA 2 Často je přínosem vědět, co se stane se vzorkem, jestliže nejsou dodrženy určené podmínky přepravy. Taková znalost dovoluje výrobci podávat lepší rady a z perspektivy uživatele, poskytovat lepší produkt. Hodnocení stability se může tedy provádět tak, aby se zhodnotila nestabilita za extrémnějších podmínek, než jsou ty, které se očekávají na základě plánovaného skladování a přepravních podmínek.

### 8.2.3 Obecné metody hodnocení stability

Hodnocení stability, jak dlouhodobé tak i přepravní, by mělo zahrnovat jedno nebo i více z následujících:

- zvážení všech fyzikálních, chemických a biologických vlastností daného materiálu, které by mohly důvodně ovlivnit stabilitu, včetně konkrétních chemických nebo biologických forem, které jsou certifikovány,
- přezkoumání publikovaných dat o stabilitě příbuzných materiálů v literatuře,
- přezkoumání hodnocení stability nebo dat o monitoringu u příbuzných materiálů,
- experimenty podle plánu, ať už v reálném čase nebo urychlené testy stability,
- experimenty k otestování vlivu různých způsobů skladování včetně neporušenosti obalu a metod stabilizace a konzervace.

Tam, kde se používá stabilizace, by měl výrobce charakterizovat metodu stabilizace a prověřit (jakoukoli z metod uvedených výše), že použitá metoda stabilizace je vhodná pro použití materiálu a účinná.<sup>[18]</sup>

Hodnocení stability by mělo zvážit potenciální vlivy opakovaného odběru podvzorků (včetně například vlivu opětovného otevírání, opětovného zmrazování nebo vlhkosti), pokud je toto výslovně dovoleno ve stanovených podmínkách pro užití. Tam, kde není proveditelné začlenit do hodnocení stability vliv opakovaného odběru podvzorků nebo opětovného použití, neměl by návod k použití dovolovat opakovaný odběr vzorků nebo opakované užití.

POZNÁMKA 1 „Příbuzné“ materiály, jak je uvedeno výše, mohou zahrnovat i materiály značně se odlišující od kandidátského referenčního materiálu, které sdílí některé, ale ne všechny důležité charakteristiky. Například stejná certifikovaná vlastnost v odlišné matici nebo rozpouštědle, stejný typ matrice, ale odlišné certifikované vlastnosti atd.

POZNÁMKA 2 V odstavcích 8.2.4 a 8.9 jsou uvedeny další opatření vztahující se k opakovanému odběru podvzorků nebo opakovanému používání.

NOTE 1 Transportation stability is often referred to as “short term stability”.

NOTE 2 It is often useful to know what might happen to the sample if the intended transport conditions are not maintained. This knowledge allows the producer to give better advice and, from the perspective of the user, a better product. Stability assessment can therefore be conducted to assess instability under more extreme conditions than those expected on the basis of planned storage and transport conditions.

### 8.2.3 General methods for assessment of stability

For both long-term and transportation stability, assessment of stability should include one or more of the following:

- consideration of all the physical, chemical and biological properties of the material that might reasonably affect stability, including the particular chemical or biological species certified;
- review of published literature data on stability of related materials;
- review of stability assessment or monitoring data on related materials;
- planned experiments, whether real-time or accelerated stability tests; or
- experiments to test the effect of different storage arrangements, including container integrity and stabilization or preservation methods.

Where stabilization is used, the producer should define the method of stabilization and verify (by any of the methods above) that the stabilization method used is appropriate for the use of the material and effective.<sup>[18]</sup>

Stability assessment should consider potential effects of repeated sub-sampling (including, for example, the effects of reopening, re-freezing, or humidity), when this is explicitly permitted under stated conditions for use. Where it is not feasible to include the effect of repeated sub-sampling or re-use in the stability assessment, the instructions for use should not permit re-sampling or re-use.

NOTE 1 “Related” materials above can include appreciably different materials from the candidate reference material that share some but not all relevant characteristics. For example, the same certified property in a different matrix or solvent, the same matrix type but different certified properties, etc.

NOTE 2 Subclauses 8.2.4 and 8.9 include further provisions related to repeated sub-sampling or repeated use.

## 8.2.4 Potřeba experimentální studie stability

### 8.2.4.1 Obecné požadavky pro potřebu experimentální studie stability

U referenčních materiálů, se kterými má výrobce málo zkušeností se stabilitou za plánovaných podmínek skladování a/nebo přepravy, nebo žádnou předchozí zkušenost, nebo tam, kde nejsou známy vlivy povoleného opakovaného odběru vzorků nebo opětovného použití, se obvykle vyžaduje experimentální hodnocení stability.

Experimentální studie nejsou nutností, pokud má výrobce předchozí zkušenosti se stabilitou velmi podobných materiálů držných po dlouhou dobu při stejných plánovaných podmínkách skladování. „Úzce podobné“ materiály jsou materiály charakterizované pro stejné vlastnosti, které mají stejné složení matrice, podmínky zpracování, podobné nebo méně efektivní balení, atd.

Odstavce 8.3 až 8.8 obsahují návod na formu a provedení experimentální studie stability.

### 8.2.4.2 Potřeba experimentální studie stability u výroby opakovaných šarží

Když

- experimentální studie je považována za nutnou a
- dotyčný materiál se vyrábí v opakovaných šaržích a
- výrobce zamýšlí používat data o stabilitě z jedné šarže k popisu stability následujících šarží,

výrobce by měl obvykle prověřit stabilitu nejméně dvou šarží experimentálně kromě šarže zvolené pro experimentální charakterizaci stability

## 8.3 Klasifikace studií stability

### 8.3.1 Obecně

Přístupy ke studii stability se mohou významně lišit. Nejjednodušší přístup spočívá v provedení dvou nezávislých měření v různých časových úsecích, aby se dal vyvodit závěr o budoucí stabilitě na základě změny v uplynulé době. Dalšími přístupy jsou přístupy na mnoha časových bodech, které umožňují výrobcí referenčního materiálu modelovat komplexnější projevy stability a/nebo činit závěry o změnách v kratších obdobích než je doba, která uplynula mezi prvním a posledním měřením studie stability.

Studie stability lze třídit na

- klasické a isochronní studie, podle podmínek měření,

## 8.2.4 Need for experimental study of stability

### 8.2.4.1 General requirements on the need for experimental study of stability

For reference materials for which the producer has little or no prior experience of stability under the planned storage and/or transport conditions, or where the effects of permitted re-sampling or re-use are not known, experimental assessment of stability is usually required.

Experimental studies are not necessary if the producer has prior experience of stability from closely similar materials held for an extended period under the same planned storage conditions. “Closely similar” materials are materials characterized for the same properties, which share the same matrix composition, processing conditions, similar or less effective packaging, etc.

Subclauses 8.3 to 8.8 give guidance on the form and conduct of experimental stability studies.

### 8.2.4.2 Need for experimental study of stability in repeated batch production

When

- an experimental study is considered necessary and
- the material in question is to be produced in repeated batches and
- the producer intends to use the stability data from one batch to represent the stability of the following batches,

the producer should normally verify the stability of at least two batches experimentally in addition to the batch chosen for experimental characterization of stability.

## 8.3 Classification of stability studies

### 8.3.1 General

Approaches to a stability study can vary significantly. The simplest approach is to take two independent measurements at different points in time, in order to draw conclusions about future stability based on change over the elapsed time. Other approaches include multi-point approaches, which enable the reference material producer to model more complex stability behaviour and/or draw conclusions about change over shorter periods than the time elapsed between the first and last measurement in the stability study.

Stability studies can be classified into

- classical and isochronous studies, according to the conditions of measurement,



- studie v reálném čase a zrychlené studie podle použitých podmínek a zpracování a
- přepravní a dlouhodobé studie podle toho, zda se zaměřují na vyhodnocení stability při přepravě nebo na stabilitu při skladování.

Kromě toho může být nutné zkoumat konkrétní podmínky, jako jsou vlastnosti balení nebo vlhkost, a ty vyžadují speciální návrhy experimentů.

V následujících odstavcích se uvádí návod pro tyto různé možnosti volby návrhů experimentů.

### 8.3.2 Klasifikace podle podmínek měření

#### 8.3.2.1 Klasické studie stability – mezilehlé podmínky měření

Při klasické studii stability se připravují vzorky současně (tj. jako šarže) a měří se za identických podmínek, až čas uběhne (tj. jeden vzorek okamžitě, další po třech měsících, další po šesti měsících, atd.). Tento návrh experimentů, ve kterém se provádějí měření za mezilehlých podmínek měření (někdy nazývaných podmínkami vnitrolaboratorní reprodukovatelnosti) mohou vést k relativně velké nejistotě, pokud nestabilita měřicího systému přispívá významnou měrou k disperzi výsledků měření.

#### 8.3.2.2 Izochronní studie stability – měření za podmínek opakovatelnosti

Izochronní návrh experimentů používá skladování za referenčních podmínek. Referenční podmínky představují soubor podmínek, za kterých lze oprávněně předpokládat, že předmětné vlastnosti zůstávají stálé. Může též být zvolena základní úroveň pro provedení měření jednotek, které jsou vystaveny různým degradačním podmínkám, ale pak změřeny v krátké době, ideálně za podmínek opakovatelnosti.<sup>[19]</sup> Slovo „izochronní“ zdůrazňuje, že měření jsou provedena současně spíše, než rozdělena do časového období po které trvá studie stability, jako je tomu u klasického přístupu. Očekává se, že toto použití podmínek opakovatelnosti povede ke zlepšené preciznosti měření během studie stability a tak lepší výkon studie stability. Následkem toho izochronní studie stability teoreticky vede k menší nejistotě než klasická studie v závislosti na rozdílu mezi opakovatelností a (vnitrolaboratorní) reprodukovatelností měření. Nezbytným předpokladem pro tento návrh je to, že mohou být určeny podmínky, za kterých nedochází k degradaci, nebo přinejmenším ta postupuje s odlišnou rychlostí než u podmínek zvolených pro skladování.

Podrobnější popis izochronních studií je uveden v odstavci 8.4.3.

- real-time studies and accelerated studies, according to the conditions and treatments applied, and
- transportation and long-term stability studies, according to whether they are aimed at evaluating stability under transport conditions or long-term stability in storage.

In addition, it may be necessary to examine particular conditions, such as packaging features or humidity, and this may require special experimental designs.

The following subclauses provide guidance on these different design options.

### 8.3.2 Classification according to conditions of measurement

#### 8.3.2.1 Classical stability studies - Intermediate conditions of measurement

In the classical stability study, individual samples prepared at the same time (i.e. as a batch), under identical conditions, are measured as time elapses (e.g. one sample immediately, one after three months, the next one after six months, etc.). This design, in which the measurements are carried out under intermediate conditions of measurement (sometimes called within-laboratory reproducibility conditions), can lead to a relatively high uncertainty when instability of the measurement system contributes significantly to the dispersion of the measurement results.

#### 8.3.2.2 Isochronous stability studies - Repeatability conditions of measurement

Isochronous designs use storage under reference conditions. Reference conditions are a set of conditions under which the properties of interest can be reliably expected to be stable. It can also be a chosen baseline level to allow measurements on units exposed to different degradation conditions to be measured in a short period of time, ideally under repeatability conditions.<sup>[19]</sup> The word “isochronous” emphasizes that the measurements are made at the same time, rather than distributed over the time span of the stability study, as is the case in the classical approach. This use of repeatability conditions is expected to lead to improved measurement precision over the course of the study, thus improving the power of the stability study. As a consequence, the isochronous stability study, in theory, leads to a smaller uncertainty than that of the classical study, depending on the difference between the repeatability and the (within-laboratory) reproducibility of the measurements. A prerequisite for this design is that conditions can be defined under which degradation does not occur, or at least occurs at a different rate from the conditions selected for storage.

A more detailed description of isochronous studies is given in 8.4.3.

### 8.3.3 Klasifikace podle trvání a podmínek studie stability

#### 8.3.3.1 Studie v reálném čase

U studií stability v reálném čase se studuje stabilita materiálu za podmínek skladování nebo podmínek přepravy, které jsou určeny pro daný RM. To znamená, že za jeden týden/měsíc/rok studie stability se získá informace o chování materiálu v průběhu týdenní/měsíční/roční doby. Tento typ studie má výhodu, že nevyžaduje žádné předpoklady o vlivu jednotlivých podmínek na stabilitu, protože podmínky, použité v experimentální studii se shodují s těmi určenými pro použití při přepravě a skladování.

#### 8.3.3.2 Urychlené studie stability

U urychlených studií stability jsou prováděny násobné experimenty za podmínek, které jsou extrémnější než podmínky skladování nebo přepravy určené pro RM s cílem vyvolat rychlejší degradaci než tu, která by se očekávala při plánovaných podmínkách skladování. Rychlost degradace za předemných podmínek se pak odhadne například regresí za různých experimentálních podmínek. Nejčastějším příkladem je testování při několika teplotách a odhad rychlosti degradace pomocí Arrheniovy rovnice. Tento druh studie má za výhodu zkrácení celkové potřebné doby a zvýšenou jistotu použitím informací z vystavení extrémnějším podmínkám. Největší nevýhodou je, že mechanismus degradace nebo stupeň určující její rychlost se může měnit v závislosti na různých podmínkách při studii, zejména při různých teplotách.

*POZNÁMKA 1* Jak studie v reálném čase, tak i urychlené studie se mohou organizovat jako klasické nebo jako izochronní studie.

*POZNÁMKA 2* Použití urychlených studií může poskytovat důvěru ve stabilitu po období podstatně delší než je doba trvání studie a je zvláště užitečné, když se vyžaduje brzká dostupnost daného materiálu.

*POZNÁMKA 3* Pokyn ISO 34<sup>[2]</sup> uvádí světlo, vlhkost a teplotu jako příklady faktorů, které mohou urychlovat degradaci.

### 8.3.4 Klasifikace podle cíle studie

#### 8.3.4.1 Studie přepravní nebo jiné krátkodobé stability

Přepravní stabilita je vlastností materiálu, která vypovídá o stabilitě za očekávaných podmínek přepravy. Za tímto účelem se studuje chování materiálu a jeho hodnot vlastností za (přinejmenším) plánovaných podmínek balení a přepravy. Čím restriktivnější tyto podmínky jsou, tím užší je rozsah podmínek, které je třeba začlenit do studie přepravní stability.

### 8.3.3 Classification according to stability study duration and conditions

#### 8.3.3.1 Real-time stability studies

In a real time stability study, the stability of a material is studied under the storage or transport conditions that are intended for the RM. This means that one week/month/year of the stability study gives information on the behaviour of the material over a one week/month/year period. This type of study has the advantage that it does not require any assumptions about the effects of different conditions on the stability, because the conditions used in the experimental study are the same as those intended for use in transport or storage.

#### 8.3.3.2 Accelerated stability studies

In an accelerated study, multiple experiments are performed at conditions that are more extreme than the storage or transport conditions intended for the RM and that aim at inducing more rapid degradation than would be experienced under the intended storage conditions. The degradation rate at the conditions of interest is then estimated, for example, by regression over the various experimental conditions. The most frequent example is testing at several temperatures and estimating a degradation rate via the Arrhenius equation. This type of study has the advantages of reducing the total time required and of increased confidence from the use of information from more extreme exposure conditions. The major disadvantage is that the degradation mechanism or its rate-determining step may change with different conditions under study, particularly the temperature.

*NOTE 1* Both real-time and accelerated studies can be organized as classical or as isochronous studies.

*NOTE 2* The use of accelerated studies can provide confidence in stability for periods substantially longer than the study duration and is particularly useful where early availability of the material is required.

*NOTE 3* ISO Guide 34<sup>[2]</sup> includes light, moisture and temperature as examples of factors that may accelerate degradation.

### 8.3.4 Classification by study objective

#### 8.3.4.1 Transportation or other short-term stability studies

Transportation stability is a property of the material referring to stability under expected transport conditions. For this, the behaviour of the material and its property values are studied under (as a minimum) the intended conditions for packaging and transport. The more restrictive the intended conditions for transport are, the smaller is the range of conditions that

Tam, kde není žádná zkušenost s konkrétním typem materiálu, měla by se studie přepravní stability uskutečnit, aby se získaly informace týkající se vhodných podmínek pro přepravu. Trvání a podmínky, které se zahrnou do studie přepravní stability, by měly odrážet trvání a podmínky, které se dají oprávněně očekávat při přepravě jednotky RM do prostor uživatele. Pokud neexistuje důvod k opaku, tyto podmínky by měly zahrnout extrémní teploty, které se mohou oprávněně vyskytnout v mezinárodní přepravě po dobu, která přinejmenším tak dlouhá jak je dovoleno pro přepravu. Například jestliže doba přepravy je limitována na tři týdny, studie krátkodobé stability na 3-4 týdny bude postačovat.

Pokud již existují zkušenosti se stabilitou podobných materiálů a ukazuje se, že budou vyžadovány přísně kontrolované podmínky přepravy (například přeprava v suchém ledu nebo ledových zábalech) může se studie omezit na plánované podmínky přepravy.

Další krátkodobé studie mohou být zapotřebí pro stanovení očekávaných podmínek skladování v prostorách koncových uživatelů. Studie, které by stanovily vhodnost alternativního způsobu skladování po krátká období v prostorách koncových uživatelů, by měly vycházet z podmínek, které dovoluje návod k použití (viz 8.9). Za tímto účelem není obvykle nutné zkoumat přísnější podmínky.

**POZNÁMKA** V tomto pokynu se předpokládá, že podmínky přepravy budou obvykle pod bedlivou kontrolou, aby vliv přepravy významně neovlivnil hodnoty vlastností a nevyžadoval zvýšení nejistoty certifikovaných hodnot. Předpokládá se tedy, že studie přepravní stability budou navrhovány tak, aby zkontrolovaly, zda nedochází k významným změnám spíše, než aby poskytly kvantitativní odhad velikosti změny.

### 8.3.4.2 Studie dlouhodobé stability

Studie dlouhodobé stability se provádějí pro zhodnocení stability za podmínek skladování specifikovaných pro dobu životnosti produktu. Dlouhodobé studie v reálném čase trvají obvykle 12 měsíců nebo více, urychlené studie jsou obvykle kratší, ale používají extrémnější podmínky. Doba platnosti certifikátu je též zajištěna monitorováním stability po vydání. Obecně, čím méně údajů je dostupných o stabilitě hodnoty vlastnosti v materiálu, tím rozsáhlejší musí být monitorování dlouhodobé stability po certifikaci. Tam, kde je doba uvedení nového materiálu na trh kritická, je možné omezit dlouhodobou studii stability na méně než 12 měsíců a provádět časté monitorování, aby se doplnily omezené údaje dostupné před certifikací.

### 8.3.5 Návrhy různých podmínek skladování a zacházení

Někdy může být potřebné prozkoumat vlivy více než jedné podmínek skladování a zacházení. Například

needs to be included in the transportation stability study.

When no previous experience is available concerning a particular type of material, a transportation stability study should be carried out to gain information concerning the appropriate conditions for transport. The duration of, and conditions included in a transportation stability study should reflect the duration and conditions reasonably expected in transporting a unit of the RM to the user's premises. Unless there is reason to the contrary, these conditions should include extreme temperatures that might reasonably occur during international transport for a period that is at least as long as that allowed for transport of the RM. For example, if the proposed transport time is restricted to 3 weeks, a short-term stability study of 3 to 4 weeks will suffice.

If experience of the stability of similar materials is available and indicates that closely controlled transport conditions are required (for example, transportation packed with dry ice or ice-packs), the study may be restricted to the intended conditions of transport.

Other short-term studies may be necessary, for example, to establish the suitability of expected storage conditions at the end user's premises. Studies to establish the suitability of alternative storage for short periods at the end user's premises should reflect the conditions permitted by the instructions for use (see 8.9). It is not usually necessary to explore more stringent conditions for this purpose.

**NOTE** It is assumed in this Guide that transport conditions will normally be sufficiently closely controlled that the effect of transport does not significantly affect the property values and does not require an increase in uncertainty for certified values. Transportation stability studies are therefore assumed to be designed primarily to check for significant change rather than to provide an accurate quantitative estimate of change.

### 8.3.4.2 Long-term stability studies

Long-term stability studies are conducted to assess stability under storage conditions specified for the life time of the product. Real-time long-term studies typically last 12 months or more; accelerated studies are typically shorter but include more extreme conditions. The period of validity of the certificate is also ensured by stability monitoring after release. In general, the less data are available about the stability of a property value in a material, the more extensive the long-term stability and post certification monitoring has to be. Where time-to-market for new materials is crucial, it is possible to limit the long-term stability study to less than 12 months and perform frequent monitoring to complement the limited data available before certification.

### 8.3.5 Designs for different storage and treatment conditions

It may sometimes be necessary to examine the effects of more than one storage or treatment condition. For

může být potřeba zhodnotit vlivy vysoké vlhkosti, vysokých a nízkých teplot, různých metod stabilizace, vliv vystavení světlu nebo různých variant balení.

Mezi experimenty pro tento účel patří:

- individuální studie každého možného vlivu na stabilitu,
- multifaktorové experimenty určené k vyhodnocení vlivu několika skladovacích podmínek a/nebo variant balení simultánně,
- studie ‚nejhoršího případu‘, které vystavují materiál té nejextrémnější kombinaci podmínek, na základě toho, že stabilita za nejextrémnějších podmínek poskytne důkaz stability za méně extrémních souborů podmínek.

Tam, kde se musí studovat mnohočetné skladovací a balicí podmínky doporučuje se použití návrhů pro faktoriální experimenty nebo dílčí faktoriální experimenty z důvodů efektivity. Tam, kde se použijí takové experimenty, by měly být možné interakce mezi různými faktory. Dvoufaktorové interakce, jako je kombinovaný vliv vlhkosti a teploty, jsou často významné a návrh experimentů by měl umožňovat je odhadnout nebo otestovat, kde je to možné.

**POZNÁMKA 1** Poučení o těchto faktoriálních nebo dílčích faktoriálních návrzích experimentů lze nalézt v odkazu <sup>[11]</sup> a jejich použití při testování stability (léčivých látek a přípravků) popisuje odkaz <sup>[20]</sup>.

**POZNÁMKA 2** Tam, kde se v multifaktoriálním experimentu prokáže vliv, je důležité ověřit vliv další studií.

**POZNÁMKA 3** V literatuře o návrzích experimentů dílčí faktoriální návrhy, umožňující odhad hlavních vlivů bez vychýlení z dvoufaktorových interakcí se někdy uvádějí jako návrhy typu ‚rozišení IV‘. Návrhy ‚rozišení V‘ pak navíc umožňují oddělení dvoufaktorových interakcí navzájem.

## Obecné požadavky na efektivní studie stability

### 8.4.1 Přehled požadavků

pro získání spolehlivých výsledků ve studii stability je důležité:

- vybrat reprezentativní podmnožinu materiálu,
- zvolit vhodnou měřicí metodu o vyhovující preciznosti a selektivitě,
- provést měření za vhodných podmínek podle vhodného návrhu experimentů a
- udělat statistickou analýzu pomocí platných statistických metod.

V odstavcích 8.4.2 až 8.4.4 je uveden návod pro první tři z těchto bodů. Statistickou analýzou se podrobně zabývá odstavec 8.5.

example, it may be necessary to assess effects of high humidity, high and low temperature, different stabilization methods, effect of exposure to light, or different packaging options.

Experiments for this purpose may include:

- individual study of each possible effect on stability;
- multi-factor experiments intended to evaluate the effect of several storage and/or packaging treatments simultaneously;
- ‘worst case’ studies that expose a material to the most extreme combination of conditions, on the basis that stability under the most extreme conditions provides evidence of stability under less extreme sets of conditions.

Where multiple storage and/or packaging conditions need to be studied, the use of factorial or fractional factorial designs is encouraged for efficiency reasons. Where used, such experiments should allow for the possibility of interaction between different factors. Two-factor interactions, such as the combined effect of humidity and temperature, are often important and the experimental design should allow for estimation or testing of these, where practicable.

**NOTE 1** Guidance on these factorial and fractional factorial experimental designs can be found in Reference <sup>[11]</sup> and their use for stability testing (of drug substances and products) is described in Reference <sup>[20]</sup>.

**NOTE 2** Where an effect is found in a multi-factor experiment, it is important to verify the effect by further study.

**NOTE 3** In the experimental design literature, fractional factorial designs that allow for the estimation of main effects without bias from two-factor interactions are sometimes known as ‘resolution IV’ designs. Resolution V designs additionally allow separation of two-factor interactions from other two-factor interactions.

## 8.4 General requirements for effective stability studies

### 8.4.1 Overview of requirements

To obtain reliable results in a stability study, it is important to:

- select a representative subset of material,
- choose a suitable measurement method with sufficient precision and selectivity,
- make the measurements under suitable conditions following an appropriate experimental design, and
- conduct the statistical analysis using valid statistical methods.

Subclauses 8.4.2 to 8.4.4 provide guidance on the first three of these points. Statistical analysis is considered in detail in 8.5.

### 8.4.2 Výběr jednotek

Jednotky pro studii stability připravených a balených materiálů se obvykle vybírají náhodně ze souboru balených jednotek

Předběžnou studii stability je však možno provést na menších šaržích materiálu, aby se stanovilo vhodné balení nebo předběžná úprava. V takových případech postačí připravit jednotky speciálně pro předběžné studie a podrobit studiu většinou množinu toho, co se tak vyrobilo.

**POZNÁMKA 1** Interpretace studie stability je obecně méně závislá na náhodném výběru jednotek, než jsou studie homogenity nebo přidělení hodnot. Užitečné předběžné informace o stabilitě lze pravděpodobně získat na základě jednoduchých strategií odběru vzorků jako je systematické vzorkování. Je však důležité se vyhnout schématu vzorkování, které by poskytlo nerepresentativní informace o stabilitě. Například výběr jednotek z vnějších partií velkého souboru jednotek RM – jako z vrchu nebo po straně velké nádoby nebo skladovací partie – může mít za následek výběr materiálu exponovaného světlu nebo výraznějším změnám teploty. Takové jednotky mohly již podlehnout nějaké změně a mohou se chovat nerepresentativně při dalším hodnocení stability, zvláště podléhá-li materiál rychlé prvotní změně a vykazuje následně přijatelnou stabilitu.

**POZNÁMKA 2** Ustanovení 7.4.3 poskytují příklady vhodných schémat vzorkování pro výběr jednotek RM pro studii.

### 8.4.3 Vhodná měřicí metoda (metody) pro studie stability

Nezávisle na návrhu studie má výstup smysl jen tehdy, je-li směrodatná odchylka výsledků měření dostatečně malá za celé období studie, eventuálně ve spojení s homogenitou mezi jednotkami.

U isochronních studií (viz níže) nebo jiných návrhů experimentů, které spoléhají především na dobrou opakovatelnost, by se měřicí metoda (metody) měly volit především se zřetelem na dobrou opakovatelnost. U studií, které mohou být ovlivněny variabilitou mezi sériemi měření, jako je jednoduchý klasický návrh při jedné podmínce skladování, by se měly metody vybírat se zřetelem na dobrou mezilehlou preciznost.

### 8.4.4 Vhodný návrh experimentů

Paletu návrhů pro experimentální studii stability uvádí odstavec 8.3. Mohou se použít i alternativní návrhy tam, kde se prokáže jejich efektivita.

Stupeň replikace u studie stability a trvání studie závisí na mnoha faktorech. Především nejistoty v odhadované degradaci závisí na trvání studie a počtu replikátů. Při rozhodování o počtu replikátů, počtu odlišných expozičních v čase a počtu a rozsahu různých podmínek expozičních, by se mělo zvažovat:

### 8.4.2 Selection of units

Units for a stability study on prepared and packaged materials should normally be selected randomly from the set of packaged units.

Preliminary stability studies may, however, be conducted on smaller batches of material in order to establish appropriate packaging or pre-treatment. In such cases, it is sufficient to prepare units specifically for preliminary studies and to study any majority subset of those so prepared.

**NOTE 1** Stability study interpretation is generally less dependent upon random selection of units than are homogeneity studies or value assignment; useful indicative information about stability is likely to be obtained from simple sampling strategies such as systematic sampling. It is, however, important to avoid any sampling scheme which may result in unrepresentative stability information. For example, selection of units from the outer regions of a large set of RM units - such as the top or side of a large container or storage area - may result in selection of material exposed to light or more extreme temperature variation. Such units may have already undergone some change and may behave unrepresentatively in further stability assessments, particularly if the material is susceptible to rapid early change and shows subsequent approximate stability.

**NOTE 2** The provisions of 7.4.3 provide examples of appropriate sampling schemes for selecting RM units for study.

### 8.4.3 Suitable measurement method(s) for stability studies

Irrespective of the study design, the outcome is only meaningful if the standard deviation of measurement results over the study time scale, possibly in conjunction with the between-unit homogeneity, is sufficiently small.

For isochronous studies (below) or other designs which rely primarily on good repeatability, the measurement method(s) used should be selected primarily for good repeatability. For studies which may be affected by run-to-run variations in measurement, such as the simple classical design at a single storage condition, methods should be selected primarily for good intermediate precision.

### 8.4.4 Appropriate experimental design

A range of designs for the experimental study of stability is given in 8.3. Alternative designs may be used where their effectiveness can be demonstrated.

The degree of replication required in a stability study and the duration of the study depend on a variety of factors. In particular, uncertainties in predicted degradation depend both on the study duration and the number of replicates. In deciding the number of replicates, the number of different exposure times, and the number and range of different sets of exposure conditions, consideration should be given to:

- preciznost, která je k dispozici u metody měření vybrané pro studii; metody s velkou precizností vyžadují nízkou úroveň replikace,
- potřeba počítat s nedostatky v individuálních pozorováních nebo jednotkách RM. Zvolit alespoň dvě jednotky RM na každou kombinaci času a teploty (nebo jiných podmínek) což poskytne rezervu a pomůže se vyhnout tomu, aby některé soubory podmínek chyběly,
- plánovaná životnost a frekvence monitorování u materiálu. Dlouhé intervaly monitorování vyžadují buď zdlouhavé studie stability, nebo četnou replikaci pro to, aby poskytly spolehlivé předpovědi degradace.
- potřeba ověření lineárních (či jiných) modelů chování. Aby se realizovala alespoň nezbytná kontrola linearitu je třeba minimálně tři časových bodů pozorování u izochronní studie, u neizochronních studií, kde mohou být důležité vlivy série, se vyžadují minimálně 4 časové body s replikací,
- potřeba pozorovat materiál za navrhaných podmínek skladování. Do studií by se mělo začlenit (ale nemusí se omezit jen na to) vystavení materiálu navrhaným podmínkám skladování.

U materiálů, u kterých je dostupné jen málo předchozích informací, by se měl počet použitých jednotek a replikátů volit, kromě zvažování výše uvedeného tak, aby postačoval k poskytnutí malých nejistot pro odhadovanou degradaci.

Tam, kde dostupné měřicí metody nemají dostatečnou preciznost pro spolehlivé určení stability, měly by se začlenit dodatečné replikáty pro každou z kombinací bodů v čase a v podmínkách a/nebo dodatečné body v čase. Způsob replikací, kterým je dáována přednost, závisí na základních zdrojích variability, jak je uvedeno dále

- Tam, kde opakovatelnost měření je hlavním zdrojem variability by se měl zvýšit počet opakovaných měření každé jednotky a/nebo počet studovaných jednotek pro každou vyhodnocovanou kombinaci času a jiných podmínek.
  - Tam, kde heterogenita RM (vyjádřená směrodatnou odchylkou mezi jednotkami  $s_{bb}$ ) je důležitým zdrojem variability by se měl zvýšit počet studovaných jednotek při každé vyhodnocované kombinaci čas/podmínky.
  - Tam, kde variabilita měření s časem je u klasické studie stability významná, by se měl zvýšit počet vyhodnocovaných časových bodů.
- the precision available from the measurement method chosen for the study; very precise methods require low levels of replication;
  - the need to allow for failures in individual observations or RM units. Allowing a minimum of two RM units for each combination of time and temperature (or other conditions) provides for redundancy and helps to avoid missing sets of conditions;
  - the planned lifetime and monitoring frequency for the material. Long monitoring intervals require either lengthy stability studies or higher replication to provide reliable predictions of degradation;
  - the need to verify linear (or other) model behaviour. To provide some check on linearity, a minimum of three observation times is essential in an isochronous study; for non-isochronous studies, where run effects may be important, a minimum of four points in time, with replication, is required;
  - the need to observe the material under proposed storage conditions. Studies should include (but need not be limited to) exposure under the proposed conditions of storage.

For materials with little prior information available, the number of units and replicates used should, in addition to the considerations above, be sufficient to provide small uncertainties for predicted degradation.

Where the available measurement methods do not have sufficient precision for reliable determination of stability, additional replicates at each combination of point in time and conditions and/or additional points in time should be included. The preferred nature of replication depends on the principal sources of variation, as follows.

- Where the measurement repeatability is the principal source of variation, the number of replicated measurements on each unit and/or the number of units studied at each combination of point in time and other conditions should be increased.
- Where RM heterogeneity (represented by the between-unit standard deviation  $s_{bb}$ ) is an important source of variation, the number of units studied at each time/condition combination should be increased.
- Where measurement variation over time in a classical stability study is important, the number of points in time should be increased.

## 8.5 Vyhodnocení výsledků studie stability

### 8.5.1 Obecné úvahy o zpracování dat studie stability

Zpracování dat ze studií stability by mělo brát v úvahu konkrétní cíle studie, použitý návrh experimentů a zdroje variability, které mohou ovlivňovat výsledky.

U většiny základních studií stability je cílem buď testování na významné změny v čase při skladování, nebo odhad rychlosti změny hodnot vlastností s časem. U studií, které vyhodnocují vliv jedné nebo různých podmínek skladování (8.3.5 výše), analýza dat je obvykle testováním významných vlivů. Některé návrhy obsahují prvky zaměřující se na oba tyto cíle.

Použitý návrh experimentů ovlivňuje možnosti analýzy dat. Návrh omezený na dva soubory podmínek – například dva extrémní body v čase – se může hodnotit jako testování významné změny mezi těmito skupinami nebo (za předpokladu opakování u každého bodu) zkoumáním regresí a v tomto jednoduchém případě jsou oba přístupy ekvivalentní. Návrh, ve kterém se sleduje hodnota vlastnosti v čase za jednoho souboru podmínek skladování, se hodnotí obvykle pomocí modelu (jako je lineární regrese, níže) pro odhad rychlosti změny nebo analýzou rozptylu pro prověрку významných změn v čase, ale nemůže být hodnocen jednoduchým dvouskupinovým testováním. Urychlené studie, ve kterých se sleduje materiál v čase při několika podmínkách skladování, mohou vyžadovat obecnější model, který by dovoľoval odhad odlišných rychlostí změn při různých podmínkách (například teplotní závislost rychlostí degradace).

Zdroje náhodné variability, které mají vliv na výsledek, též ovlivňují způsob analýzy dat. U nejjednodušších návrhů, ve kterých je přítomen jen jeden zdroj náhodné variability, obvykle postačuje jednoduchá regrese popsaná v odstavci 8.5.2. U studií s měřeními v různém čase a ve kterých může měřicí systém vykazovat variabilitu mezi sériemi kromě variability v sérii, měla by se volit analýza dat tak, aby dovolila začlenit i další zdroj (zdroje) náhodné variability. Analýzou dat pro takové případy se zabývá odstavec 8.5.5.

Následující odstavce popisují analýzu dat studie stability. V odstavci 8.5.2 se popisuje aplikace lineární regrese pro nejjednodušší případ. Dále je poskytnut návod pro komplexnější případy včetně isochronních studií za jedné podmínky skladování a za použití jedné referenční podmínky, urychlených studií a studií s komplexnější chybovou strukturou.

**POZNÁMKA** Vyhodnocení nejistoty referenčních materiálů může též využívat výsledky studií stability.

## 8.5 Evaluation of stability study results

### 8.5.1 General considerations for stability study data treatment

Data treatment for stability studies should take account of the particular study objective, the experimental design used, and the sources of variation that may affect the results.

For most basic stability studies, the objective is either to test for any important change over time in storage or to estimate the rate of change of property values over time. In studies that examine the effect of single or multiple storage conditions (8.3.5 above), data analysis is normally intended to provide a test for significant effects. Some designs include elements of both of these objectives.

The experimental design used will affect the data analysis options. A design restricted to two sets of conditions - for example, two extreme points in time - can be assessed using either a test for significant change between two groups, or (assuming replication at each point) by regression, and in this simple case the two procedures are equivalent. A design that follows a property value over time at a single set of storage conditions is typically assessed using modelling (such as linear regression, below) to estimate rates of change or by analysis of variance to test for significant differences over time, but cannot be assessed by a simple two-group test. An accelerated study, which follows a material over time at several storage conditions, may require treatment using a more general model that allows estimation of different rates of change under different conditions (for example, the temperature dependence of rates of degradation).

The sources of random variation that affect the result also affect data analysis. For the simplest design, in which only a single source of random variation is present, simple regression as described in 8.5.2 is usually sufficient. For studies involving measurements at different times, and in which the measurement system may show run-to-run variation in addition to within-run variation, data analysis should be chosen to allow for the additional source(s) of random variation. Data analysis for such cases is considered in 8.5.5.

The following subclauses describe the analysis of stability study data. Subclause 8.5.2 describes the application of linear regression in the simplest case, and provides guidance for more complex cases including isochronous studies using one storage and one reference condition, accelerated studies, and studies with more complex error structure.

**NOTE** Uncertainty evaluation for certified reference materials can also use the results of stability studies.

## 8.5.2 Základní studie stability: několik bodů v čase při jedné podmínce skladování

### 8.5.2.1 Použitelnost

Tento odstavec se vztahuje k situaci, kdy se uskutečňují jedno či více měření při každém z časových bodů, náhodné chyby každého měření jsou nezávislé a vykazují stejnou směrodatnou odchylku (tedy chyby jsou nezávislé a mají identické rozdělení).

**POZNÁMKA 1** Předpoklad nezávislosti neplatí, pokud jsou měření provedena v různých časech a směrodatná odchylka mezilehlé preciznosti měřicího systému je vyšší než směrodatná odchylka opakovatelnosti. Viz 8.5.2.5 pro další informace.

**POZNÁMKA 2** Pokud se preciznost významně mění z jednoho časového bodu k druhému, předpoklad identického rozdělení neplatí a tento odstavec nelze aplikovat. Změny v preciznosti lze detekovat například testováním homogenity rozptylu nebo zkoumáním grafu reziduí. Viz 8.5.2.5 pro další informace.

### 8.5.2.2 Postup

Pro základní případ, používající jednoduchou lineární regresi, je postup následující:

- fitujte předběžný model a zkoumejte výsledek fitování a rezidua modelu s kontrolou stanovených předpokladů (viz 8.5.2.4 a 8.5.2.5 pro podrobnosti),
- jestliže předpoklady platí, zaznamenej parametry modelu (obvykle směrnici a úsek) a jejich nejistoty a zkontrolujte statistickou významnost zjištěných trendů (viz 8.5.2.6).

Podrobné výpočty pro jednoduchou lineární regresi uvádí dodatek B.3 v příloze B a jsou obecně součástí všech univerzálních statistických softwarových balíčků a většiny komerčních tabulkových procesorů.

**POZNÁMKA** V zásadě všechny softwary pro lineární regresi, včetně aplikací komerčních tabulkových procesorů, provádějí tyto výpočty automaticky. Software, včetně softwaru napsaného výrobcem RM pro tento účel, by se měly před použitím validovat.

### 8.5.2.3 Výběr modelu

U hodnocení stability, kde je znám výchozí kinetický mechanismus a očekávají se malé změny, je lineární aproximace obvykle vhodným modelem. Jednoduchý lineární vztah se popisuje v dodatku B.3.

V případech, kde dobře definovaný nelineární mechanismus je důvodem nestability, je třeba dávat přednost odpovídajícímu degradačnímu modelu před (empirickým) lineárním modelem. Matematické zpracování je u modelů jiných než lineárních o něco složitější, ale vyhodnocení probíhá stejným způsobem.

## 8.5.2 The basic stability study: multiple points in time at a single storage condition

### 8.5.2.1 Applicability

This subclause applies to a situation where one or more measurements are taken at each of several points in time, the random errors in each measurement are independent and share the same standard deviation (that is, errors are independent and identically distributed).

**NOTE 1** The assumption of independence is not valid when the measurements are taken at different times and the intermediate precision standard deviation of the measurement system is greater than the repeatability standard deviation. See 8.5.2.5 for further information.

**NOTE 2** If the precision changes significantly from one time to another, the assumption of identical distribution is not valid and this clause does not apply. Changes in precision may be detected by, for example, application of tests for homogeneity of variance or by inspection of residual plots. See 8.5.2.5 for further information.

### 8.5.2.2 Procedure

For the basic case, using simple linear regression, the procedure is as follows:

- fit a preliminary model and inspect the fit and model residuals, checking any assumptions made (see 8.5.2.4 and 8.5.2.5 for further detail);
- if the assumptions apply, record the model parameters (usually slope and intercept) and their uncertainties and check the statistical significance of any trend found (see 8.5.2.6).

Detailed calculations for simple linear regression are given in Clause B.3 and are generally available in all general-purpose statistical software packages and in most commercial spreadsheet applications.

**NOTE** Essentially all linear regression software, including commercial spreadsheet applications, carries out these calculations automatically. Software, including any written by the RM producer for this purpose, should be validated prior to use.

### 8.5.2.3 Model selection

For stability assessment where the underlying kinetic mechanism is unknown and changes are expected to be small, a linear approximation is usually a suitable model. The simple linear relationship is described in Clause B.3.

In cases where a well-defined nonlinear mechanism is the reason for the instability, the corresponding degradation model is to be preferred over the (empirical) linear model. The mathematics is somewhat more complex for models other than the straight line, but the evaluation runs in the same fashion.



**PŘÍKLAD** Referenční materiál, který obsahuje radioaktivní izotop je příkladem vlastnosti s dobře známým kinetickým mechanismem, v tomto případě radioaktivního rozpadu, který se může dobře předpovídat dobře známým, ale nelineárním modelem.

#### 8.5.2.4 Fitování modelu

Parametry regrese se mohou vypočítat pomocí postupů v odstavci B.3 nebo vhodným softwarem. Výpočty poskytnou odhady  $b_0$  a  $b_1$ , pravých hodnot směrnice a úseku  $\beta_0$  a  $\beta_1$  v uvedeném pořadí, spolu s odpovídajícími směrodatnými chybami  $s(b_0)$  a  $s(b_1)$ , které se mohou použít pro následné statistické testování a při vyhodnocování nejistoty.

#### 8.5.2.5 Prozkoumání a ověření předpokladů

Výsledky regrese se mají zkoumat a předpoklady regrese ověřovat tak jak je naznačeno v dodatku B.3.3.

U studií stability referenčních materiálů je zvláště důležité kontrolovat projevy svědčící o nevhodnosti degradačního modelu. V případě jednoduchého lineárního modelu, mají tudíž doklady o nelinearitě zvláštní význam.

Tam, kde se zjistí značná odchylka od předpokladů, měla by se analýza dat přerušit a anomálie řešit nebo přijmout alternativní zpracování dat.

#### PŘÍKLADY

- Odlehlé body by se měly zkoumat, a popřípadě vyloučit. Odlehlé body nepříznivě ovlivňují fitování modelu i testování významnosti. Odlehlé hodnoty, pokud je lze přiřadit k selhání měřicího systému, by se měly opravit nebo vyloučit. Avšak odlehlé body mohou mít též příčinu v nesprávné volbě modelu degradace nebo patřit jednotkám RM, které degradovaly individuálně. Je tedy třeba věnovat příslušnou péči zvážení možných příčin před tím, než jsou jakákoliv pozorování vylučována. Jednotlivá odlehlá pozorování mezi replikáty stejné jednotky RM jsou při studii stability, pokud již daný materiál prokázal, že je homogenní, pravděpodobným projevem selhání měření.
- Pokud je heterogenita mezi jednotkami významná (ve srovnání s platnou precizností měření), je nutné s tím počítat při statistickém hodnocení; zejména významnost se může jevit nadsazená, pokud se nezohlední heterogenita mezi jednotkami.
- Svědectví o variabilitě mezi sériemi v měřicím systému by se měly řešit přijetím alternativních způsobů zpracování dat (viz 8.5.5).
- Projevy zakřivení závislosti by se měly řešit použitím jiného modelu nebo (pokud viditelná změna znamená nestabilitu) učiněním závěru, že materiál není dostatečně stálý pro zamýšlené použití.

**EXAMPLE** An RM containing a radioactive isotope is an example of a property with a well-defined kinetic mechanism, in this case a radioactive decay, which can be predicted by a well-known but nonlinear model.

#### 8.5.2.4 Fitting the model

The regression parameters can be computed using the procedures in Clause B.3 or by suitable software. The calculations provide estimates  $b_0$  and  $b_1$ , respectively, of the true intercept and slope  $\beta_0$  and  $\beta_1$ , together with the corresponding standard errors  $s(b_0)$  and  $s(b_1)$ , which can be used in subsequent statistical tests and in uncertainty evaluation.

#### 8.5.2.5 Inspection and check of assumptions

The regression results should be inspected and the assumptions checked as indicated in B.3.3.

In reference material stability studies, it is particularly important to check for evidence of an inappropriate degradation model. In the case of a simple linear model, evidence of nonlinearity is therefore particularly important.

Where serious departure from the assumptions is found, data analysis should be discontinued and any anomalies should be resolved or alternative data treatment adopted.

#### EXAMPLES

- Outlying data points should be inspected, and if appropriate, removed. Outlying data points adversely affect both model fits and tests for significance. If attributable to measurement system failures, outlying data points should be corrected or removed. However, outliers may also be due to incorrect choice of degradation model or to RM units that have degraded individually; due care should accordingly be taken to consider possible causes before removing any observations. Individual outlying observations among replicates on the same unit in a stability study are, if the material has already been shown to be homogeneous, likely to indicate measurement failure.
- If there is significant between-unit heterogeneity (compared to the available measurement precision), it is necessary to allow for this in statistical assessment; in particular, significance may be over-stated if between-unit heterogeneity is not allowed for.
- Evidence of run-to-run variation in the measurement system should be addressed by adopting alternative data treatment (see 8.5.5).
- Evidence of curvature may be addressed using a different model or (since visible change implies instability) by concluding that the material is not sufficiently stable for the intended use.

### 8.5.2.6 Testování na významnou nestabilitu

U jednoduchého případu uvedeného výše, obvyklým způsobem testování nestability je použití Studentova t-testu na odlišnost směrnice od nuly, jak se popisuje v dodatku B.3.4.

Jestliže nalezený trend je statisticky významný, je též vhodné zvážit, zda je technicky významný, tedy zda postačuje požadovat zvýšení nejistoty certifikované hodnoty či zabránit certifikaci. Trendy by měly být považovány za technicky významné, jestliže předpokládaná degradace za zamýšlenou životnost materiálu je závažná ve srovnání se standardní nejistotou hodnoty předmětné vlastnosti.

Jestliže se zjistí technicky významný trend, použijí se opatření z odstavce 8.6.

### 8.5.3 Návrhy izochronních studií

#### 8.5.3.1 Jednoduchá izochronní studie s jedněmi podmínkami skladování a jednou referenční podmínkou

U izochronního návrhu experimentů je sada jednotek RM vystavována navrhovaným skladovacím podmínkám a po určité plánované době expozice každé jednotky je pak přemístěna do referenčních podmínek. Když proběhne vystavení všech jednotek po zvolené doby, měří se kompletní soubor jednotek. Tato sekvence je schematicky znázorněna na obr. 6.

Obvykle jsou plánovanými skladovacími podmínkami okolní teplota nebo teplota blízko okolní teploty (například 4°C) a referenční podmínky představují mnohem nižší teplotu, při které je degradace velmi nepravděpodobná.

Protože studie zahrnuje pouze jedny podmínky expozice, analýza dat může vycházet z postupu pro základní studii stability uvedenou v odstavci 8.5.2, a pro analýzu využít vynesení doby, po kterou je každá z jednotek vystavena plánovaným podmínkám skladování na osu času.

**POZNÁMKA** Při izochronní studii je též možné přenést materiál z referenčních podmínek do plánovaných expozičních podmínek a tak dosahovat požadovaný čas expozice.

### 8.5.2.6 Testing for significant instability

For the simple case above, the usual test for instability is a Student t-test for slope significantly different from zero as described in B.3.4.

If the trend is statistically significant, it is also useful to consider whether it is technically significant; that is, whether it is sufficient to require an increase of uncertainty in a certified value or to prevent certification. Trends should be considered technically significant if the predicted degradation over the intended lifetime of the material is important compared to the standard uncertainty of the property value in question.

If a technically significant trend is observed, the provisions of 8.6 apply.

### 8.5.3 Isochronous designs

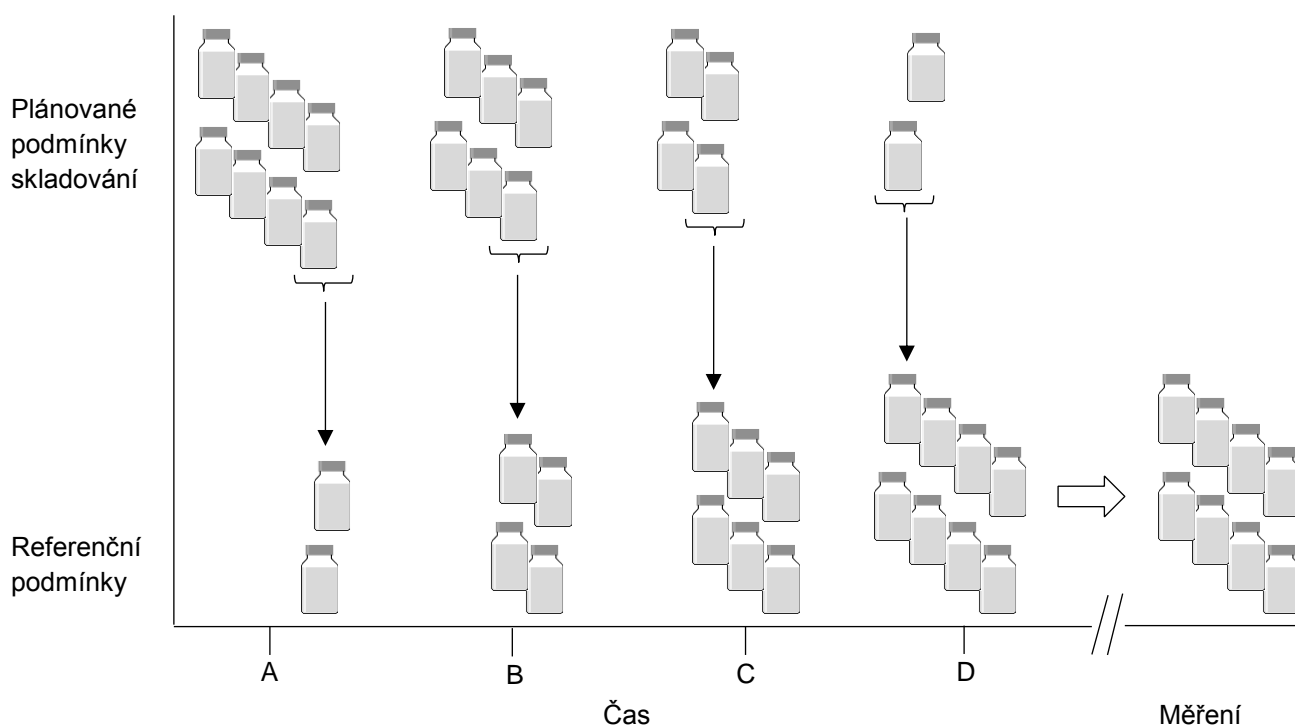
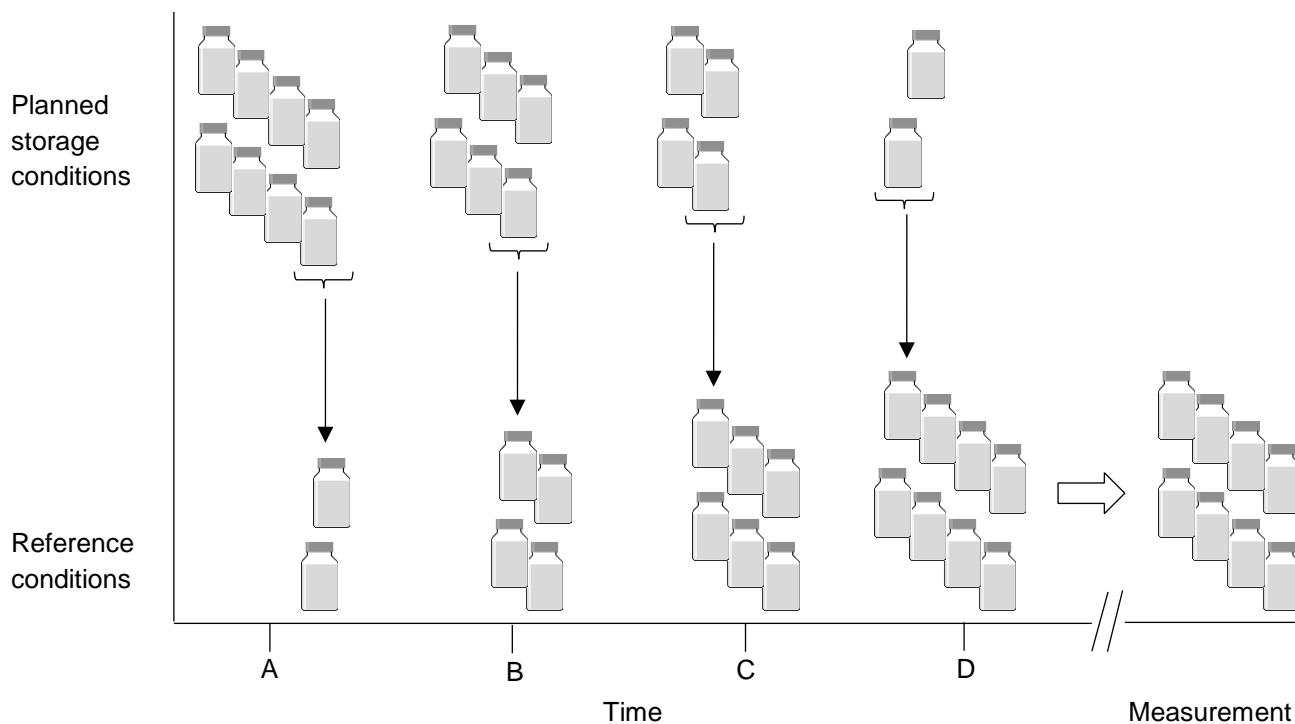
#### 8.5.3.1 Simple isochronous study with one storage and one reference condition

In an isochronous design, a set of units of an RM is exposed to proposed storage conditions for a time, and then moved to reference conditions, after the planned exposure time for each unit. When all units have been exposed for the intended time, the complete set of units is measured. This sequence is illustrated schematically in Figure 6.

Usually, the planned storage condition is at ambient or near-ambient temperature (for example, 4°C) and the reference conditions include a much lower temperature at which degradation is considered extremely unlikely.

Since the study involves only a single exposure condition, data analysis follows the procedure for the basic stability study at 8.5.2, using the time for which each unit is exposed to the planned storage conditions as the time axis for the analysis.

**NOTE** In an isochronous study, it is also possible to move material from reference conditions to planned exposure conditions to achieve the desired exposure time.



**POZNÁMKA** Na obrázku je schematicky znázorněna izochronní studie, ve které se získají pozorování po expozici ve zvyšujících se časových úsecích A, B, C, D, rovnoměrně rozvržené do zamýšlené doby studie. Počet jednotek (zde 8) je vyhrazeno pro tuto studii. Ve lhůtě A (která může být nominálním nulovým časem expozice) je jedna sada jednotek přemístěna do referenčních podmínek. Stejný počet je pak přemísťován do referenčních podmínek v následujících časových bodech. Nakonec, když všechny jednotky byly exponovány po požadované doby a pak skladovány za referenčních podmínek, se jednotky vyjmou a změří současně. Všimněme si, že měření se nemusí dělat za

**NOTE** The figure shows a schematic illustration of an isochronous study involving observations after exposure for increasing times A, B, C and D, equally spaced over the intended study period. A number of units (8, here) is reserved for the study. At time A (which may be at a nominal zero exposure time), a set of units is moved to the reference conditions. An equal number is moved to reference conditions at each subsequent point in time. Finally, after all units have been exposed for the desired time, all units are stored under reference conditions and are then removed and measured simultaneously. Note that measurements need not

referenčních podmínek, za předpokladu, že se provede v krátké době.

be conducted under the reference conditions, provided that they are carried out within a short period of time.

**Figure 6 — Illustration of an isochronous study**

**Obrázek 6 — Znárodnění izochronní studie**

### 8.5.3.2 Výhody a nevýhody izochronních studií stability

Izochronní studie, včetně jednoduchého uspořádání podle odstavce 8.5.3.1 mají následující výhody:

- Všechna měření se provádí v krátké době, obvykle za podmínek opakovatelnosti v jediné sérii. To poskytuje nejlepší možnou preciznost pro studii a zabraňuje, aby dlouhodobý drift měřicího systému mohl být považován omylem za nestabilitu.
- Završení všech měření v jedné sérii snižuje náklady a zjednodušuje naplánování laboratorních prostředků.

Mezi nevýhody patří:

- Odklad měření až na konec studie znamená, že nestabilitu nelze rozpoznat až do ukončení dané studie.
- Interpretace výsledků v základní studii podle odstavce 8.5.3.1 počítá s minimální změnou za referenčních podmínek. Pokud se materiál progresivně mění za referenčních podmínek (například kvůli progresivnímu vlivu zmrazování) může to být chybně považováno za změnu při plánovaných podmínkách skladování nebo to může vést k nesprávným závěrům, že materiál je stabilní. Je tedy třeba dodatečného důkazu pro spolehlivou interpretaci.
- Návrh studie předpokládá, že neexistuje žádný nepříznivý vliv při přenášení do nebo z referenčních podmínek. Pokud se materiál mění při změně podmínek, výsledky mohou být velmi variabilní nebo těžko interpretovatelné.
- Ne všechny materiály mohou být umístěny do podmínek, které účinněji zabraňují degradaci než plánované podmínky skladování. Některé nemohou být zchlazovány na nízkou teplotu, jiné vyžadují uchovávání při nejnižších teplotách, které jsou k dispozici.
- Ne všechny degradační mechanismy jsou ovlivňovány změnami podmínek, včetně teploty

Základní návrh podle odstavce 8.5.3.1 má další potenciální nevýhodu a to, že zahrnuje pouze jedny kandidátské podmínky skladování.

**POZNÁMKA** Nevýhody vyjmenované výše zřídka převáží výhody izochronních návrhů a mohou se dále snížit rozšířením návrhu, jak se popisuje v odstavci 8.5.3.3

### 8.5.3.2 Advantages and disadvantages of isochronous stability studies

Isochronous studies, including the simple arrangement at 8.5.3.1, have the following advantages.

- All measurements are made in a short period of time, usually under repeatability conditions in a single run. This provides for the best available precision for the study and avoids the possibility that longer-term drift in the measurement system might be mistaken for instability.
- Completing all measurements in a single run reduces costs and simplifies the scheduling of laboratory resources.

The disadvantages include:

- Deferring measurement until the end of the study means that instability is not identified until after completion of the study.
- Interpretation of results in the basic design of 8.5.3.1 relies on minimal change under the reference conditions. If the material changes progressively under the reference conditions (for example owing to progressive freezing effects), this can either be mistaken for change at the planned storage conditions or can lead to an incorrect conclusion that the material is stable. Additional evidence can therefore be needed for reliable interpretation.
- The study design assumes that there is no adverse effect on transferring to, or from, the reference conditions. If the material changes on moving between conditions the results may be highly variable or hard to interpret.
- Not all materials can be placed in conditions that are more effective at preventing degradation than the planned storage conditions. Some cannot be cooled to low temperatures; others may require storage at the lowest available temperature.
- Not all degradation mechanisms are affected by changed conditions, including temperature.

The basic design at 8.5.3.1 has the additional potential drawback of including only one candidate storage condition.

**NOTE** The disadvantages listed above rarely outweigh the advantages of isochronous designs and can be further reduced by extending the design as described in 8.5.3.3.

### 8.5.3.3 Rozšíření základního izochronního návrhu

Některé z nevýhod uvedených v odstavci 8.5.3.2 lze řešit rozšířením základního návrhu izochronního návrhu. Taková rozšíření představují

- Uskutečnění dvou studií o různé době trvání může poskytnout dříve výsledky, pokud je to žádoucí.
- Zavedení několika alternativních podmínek pro expozici, včetně 'urychlených' podmínek. To umožní hodnocení stability při různých podmínkách a může navíc umožnit izochronní měření v urychlené studii stability. Kromě toho použití dalších jednotek RM vystavených mezilehlým podmínkám (tj. mezi referenčními podmínkami a plánovanými skladovacími podmínkami) poskytne určitou zábranu proti chybné interpretaci neočekávaných změn za referenčních podmínek.
- Měření používající znáhodněné blokové návrhy. Tam, kde existuje příliš mnoho měření, které by se měly uskutečnit v jediné měřicí sérii, mohou se měření provádět s využitím znáhodněného blokového návrhu. V tomto případě vliv každého provedení expozice se projevuje v menším počtu případů u každé z několika sérií měření. To poskytuje porovnání změn ve vztahu ke stabilitě oproti směrodatné odchylce opakovatelnosti, na úkor malé ztráty ve stupních volnosti. Je to analogické jako u použití znáhodněných blokových návrhů při testování homogenity (viz 7.6.3).

Analýza dat u rozšířených návrhů pod b) a c) výše je mnohem složitější než jednoduchá lineární regrese. Na zpracování mnohateplotní izochronní studie se vztahují opatření uvedená v odstavci 8.4.4. U znáhodněných blokových návrhů jsou použité postupy podobné těm z odstavce 7.6.3 a obecně vyžadují statistický software, avšak jednoduchý průměr pozorování z každého provedení napříč sériemi měření dovolí zjednodušené hodnocení, jak je popsáno výše.

## 8.5.4 Urychlené studie stability s několikaletými podmínkami expozice

### 8.5.4.1 Popis urychlených studií

Tento typ studií urychluje vznik degradačních efektů z možných ovlivňujících faktorů, tím, že vystavuje testované předměty extrémnějším podmínkám, než podmínkám plánovaného skladování. Tak dochází ke zkrácení doby, ve které se pozorovatelná změna vyskytne <sup>[21],[22]</sup> a/nebo se zvýší pozorovatelná degradace.

Mezi faktory, které mohou urychlovat degradaci, patří, ale neomezuje se jen na ně:

- Environmentální stresové faktory jako jsou teplota, vlhkost, radiace (obvykle světlo, zvláště UV), oxidující prostředí (např. vzduch), vlastnosti obalu

### 8.5.3.3 Extensions of the basic isochronous design

Some of the disadvantages listed in 8.5.3.2 can be addressed by extending the basic isochronous design. Extensions include:

- Running two studies with different durations can provide earlier results, if desired.
- Inclusion of multiple alternative exposure conditions, including 'accelerating' conditions. This can allow assessment of stability under a range of conditions and can additionally allow isochronous measurements for an accelerated stability study. In addition, including additional RM units retained under intermediate conditions (that is, between reference and planned storage conditions) provides some protection against misinterpretation of unexpected changes under reference conditions.
- Measurement using randomized block designs. Where there are too many observations to complete in a single measurement run, the measurements can be carried out using a randomized block design. In this case, the effect of every exposure treatment is observed a smaller number of times in each of several measurement runs. This provides for a test of stability-related changes against the repeatability standard deviation, at the expense of a small loss in degrees of freedom. This is analogous to the use of randomized block designs in homogeneity testing (see 7.6.3).

Data analysis for the extended designs involved in b) and c) above is more complex than simple linear regression. For the treatment of a multiple-temperature isochronous study, the provisions of 8.4.4 apply. For randomized block designs, the procedures used are similar to those of 7.6.3 and generally require statistical software; however, a simple average of observations from each treatment across measurement runs permits simplified assessment as described above.

## 8.5.4 Accelerated stability studies with multiple exposure conditions

### 8.5.4.1 Description of accelerated studies

This type of study accelerates the occurrence of degradation effects from possible influential factors by submitting the objects under test to more extreme conditions than the planned storage condition, thus shortening the time within which an observable change occurs<sup>[21],[22]</sup> and/or increasing the observable degradation.

Factors that may accelerate degradation include, but are not limited to:

- Environmental stress factors such as temperature, humidity, irradiation (usually light, especially UV), oxidising agents (e.g. air), storage container

pro skladování (kontaminované kontaktní povrchy, adhezivní vlastnosti kontaktního povrchu, prosakování),

- Specifické faktory vzorků jako jsou reaktivní složky matrice, zejména u přírodních matricových referenčních materiálů.

Protože teplota hraje hlavní roli přinejmenším u většiny známých degradačních procesů, včetně chemických reakcí, difuze, vypařování nebo adsorpce, urychlené studie by měly obvykle testovat stresový faktor teploty. Rozsah studovaných teplot by měl zahrnovat přinejmenším rozsah teplot, se kterými by se materiál mohl pravděpodobně setkat za plánovaných podmínek skladování a přepravy.

Stresové faktory jiné než teplota, zvláště pak vlhkost a ozáření by se měly začleňovat do urychlených studií stability všude tam, kde neexistuje důkaz, že by se mohly zanedbat.

Následující paragrafy se omezují na stresový faktor teplotu, ačkoliv tento obecný přístup lze použít i na jakýkoliv další spojitý stresový faktor.

**POZNÁMKA 1** Jestliže není pozorována žádná změna při urychlené studii, pak není možné fitovat spolehlivý předpovědní model. Pro následné matematické modelování a předpovídání změn je tedy vhodné, aby se volily takové podmínky v urychlené studii, aby bylo možno pozorovat patrnou změnu alespoň při extrémech.

**POZNÁMKA 2** U dozoru nad humánními léčivými přípravky se doporučuje, aby přinejmenším teplota a vlhkost a ve speciálních případech ozáření by se měly otestovat na jejich účinek na stabilitu materiálu. Doporučení International Conference on Harmonization (ICH) popisují urychlené degradační studie, při kterých testovaný materiál je vystaven určitým definovaným kombinacím těchto faktorů.<sup>[20]</sup>

**PŘÍKLAD** Příklad možného rozvržení urychlené studie stárnutí kandidátského materiálu uvádí tabulka 1. V závislosti na očekávaném chování a rychlosti degradace, rozvržení může obsahovat jeden nebo více stresových bodů v čase a teplotě a rozsahy se mohou měnit v závislosti na druhu zkoumané měřené veličiny

properties (contaminated contact surfaces, adhesive contact surface properties, leakage);

- Sample-specific factors such as reactive matrix constituents, in particular for natural matrix RMs.

Since temperature plays a major part in at least the majority of known degradation processes, including chemical reactions, diffusion, evaporation, or adsorption, accelerated studies should normally test for the stress factor of temperature. The temperature range studied should include at least the range of temperatures the material might reasonably encounter under planned conditions of storage and transport.

Stress factors other than temperature, particularly humidity and irradiation, should be included in accelerated stability studies where there is no existing evidence that they can be neglected.

The following paragraphs are limited to the stress factor temperature, although the general approach can also be applied for any other continuous stress factor.

**NOTE 1** If no change is observed in an accelerated study then it is not possible to fit a reliable predictive model. It is therefore useful for subsequent mathematical modelling and prediction of change if the conditions included in an accelerated study can be chosen so that appreciable change is observed at least at the extremes.

**NOTE 2** For the surveillance of pharmaceutical products, it is suggested that at least temperature and humidity, and in special cases irradiation, should be tested for their impact on the stability of the material. Recommendations of the International Conference on Harmonization (ICH) describe accelerated degradation studies in which the material under test is exposed to certain, defined combinations of these factors.<sup>[20]</sup>

**EXAMPLE** An example of a possible layout of an accelerated ageing study for a candidate material is given in Table 1. Depending on the expected behaviour and the rate of degradation, a layout might contain more or less stress points in time and temperature, and ranges may vary according to the kind of measurand under investigation.

**Table 1 — Accelerated ageing example: Temperatures and exposure times**

Temperature (°C)	4	20	40	70
Exposure time (months)	1, 3, 6, 12, (24)	1, 3, 6, 12, (24)	1, 3, 6	0,5, 1
NOTE Exposure times in brackets show sampling times for the first post-certification monitoring (see 8.10)				

**Tabulka 1 — Příklad urychleného stárnutí: Teploty a doby expozice**

Teplota (°C)	4	20	40	70
Doba expozice (měsíce)	1, 3, 6, 12, (24)	1, 3, 6, 12, (24)	1, 3, 6	0,5, 1
POZNÁMKA Doba expozice uvedená v závorkách představuje dobu odběru vzorků při prvním monitoringu po certifikaci (viz 8.10)				

#### 8.5.4.2 Matematické modely pro hodnocení urychlených studií stability

Časová závislost teplotně závislé degradace předmětného analytu  $g(t)$  se nejobvykleji modeluje pomocí základního předpokladu o proporcionalitě mezi rychlostí změny analytu s časem ( $dx/dt$ ) a celkového množství analytu ve vzorku. Koeficient úměrnosti je teplotně závislá reakční rychlost  $k_{\text{eff}}(T)$ . Pak tedy z

$$\frac{dx}{dt} = -k_{\text{eff}}(T) x \quad (10)$$

kde  $x$  je hodnotou zkoumané měřené veličiny, lze získat exponenciální časovou závislost. Ačkoliv tento předpoklad není univerzálně použitelný, popisuje velký počet degradačních procesů. Jednoduché lineární modelování na základě tohoto předpokladu by mělo tedy vycházet z  $\ln[x(t)]$  spíše než z hodnot  $x(t)$  a regrese, jak jsou zmiňovány níže, by se měly provádět v log-lineárním uspořádání nebo by měly používat prokládání nelineárních křivek odpovídajících zvolenému modelu.

Při dané teplotě může vznikat efektivní reakční rychlost pseudoprvního řádu z více než jednoho souběžného degradačního procesu, z nichž každý má odlišnou závislost na teplotě. Teplotně řízené degradační procesy však často sledují Arrheniův model, biologické/mikrobiologické O'Neillův model, a změny řízené difúzí ve vzorku pak Fickův zákon. Jestliže jsou známy alespoň určité výchozí informace o možných chemických reakčních procesech, může sloužit dynamická reakční kinetika jako základ popisu závislosti reakční rychlosti na úrovni uvažovaného faktoru. Empirické modely (bez fyzikálně-chemického odůvodnění) jsou přijatelné v těch případech, kdy se překrývá velký počet jednotlivých procesů a mohou se (přínejmenším částečně) vyvažovat.

Vhodně vybraný model je pak použit pro regresi celého souboru dat ze studie stability (v návaznosti ke všem stresovým úrovním) a poskytuje pro každou stresovou úroveň odhad parametrů použitého modelu. Tyto odhady pak umožní výrobci

- odhadnout finální datum expirace materiálů s detekovanými nestabilitami (za podmínek skladování u výrobce),
- odhadnout nejistotu nestability u materiálů,
- odhadnout maximálně přípustné období pro opětovné testování u zdánlivě stabilních materiálů (monitorování po certifikaci),
- hodnotit maximálně přípustné stresové zátěže (teplotu a čas) během dodávky materiálu zákazníkovi.

Pod podmínkou v odstavci 8.4.3 a s tím, že při vhodném návrhu studie (8.4.4) se nezjistí žádná degradace na jakékoli stresové úrovni v urychlené studii, v níž nejvyšší teplota je přínejmenším 20°C nad plánovanou teplotou skladování, se pak může

#### 8.5.4.2 Mathematical models for assessing accelerated stability studies

This time dependence  $g(t)$  of the temperature-driven degradation of the analyte under consideration is most commonly modelled using the basic assumption of proportionality between the rate of change of analyte with time ( $dx/dt$ ) and the total amount of the analyte available in the sample. The proportionality coefficient is the temperature-dependent reaction rate  $k_{\text{eff}}(T)$ . Thus, from

$$\frac{dx}{dt} = -k_{\text{eff}}(T) x \quad (10)$$

with  $x$  being the value of the measurand under investigation, one obtains an exponential dependence time. Although this assumption is not universally applicable, it describes a very large number of degradation processes. Simple linear modelling following this assumption should therefore refer to  $\ln[x(t)]$  rather than the value  $x(t)$  itself, and the regressions as mentioned below should be carried out in a log-linear space or should use nonlinear curve fitting appropriate to the model chosen.

At a given temperature, the effective pseudo first-order reaction rate  $k_{\text{eff}}(T)$  may arise from more than one concurrent degradation process, each with a different dependence on temperature. Temperature-driven degradation processes however quite often follow an Arrhenius model, biological/microbial processes an O'Neill model, diffusion-driven changes in a sample Fick's law, and if at least some prior information on possible chemical reaction processes is available, the dynamic reaction kinetics may serve as the basis for describing the dependence of the reaction rates on the level of the factor. Empirical models (without a physico-chemical rationale) are acceptable in cases when a larger number of individual processes overlap and may (at least partially) compensate.

An appropriately selected model is then regressed over the whole set of data (referring to all stress levels) from the stability study providing, for any stress level, an estimate for the model parameters. These estimates then will allow the producer to

- estimate a final expiry date for materials with detected instabilities (under storage conditions at the producer's site);
- estimate an uncertainty of stability for materials;
- estimate the maximum admissible re-testing interval for seemingly stable materials (post-certification monitoring);
- assess maximum admissible stress loads (temperature and time) during delivery of the material to the customer.

Subject to 8.4.3 and to appropriate study design (8.4.4), where no degradation is found at any stress level in an accelerated study in which the highest temperature is at least 20°C above the planned storage temperature, the degradation rate under

degradační rychlost považovat za zanedbatelnou pod podmínkou, že se potvrzuje vhodnými opatřeními monitoringu (viz 8.10).

Extrapolace mimo rozsah testovaných skladovacích podmínek (například předpověď degradační rychlosti při  $-20\text{ °C}$  z experimentu při teplotách jen nad  $0\text{ °C}$ ) může být nespolehlivá a nedoporučuje se.

**POZNÁMKA 1** Procesy v pozadí degradace/zkažení materiálu mohou být složité a mohou vyžadovat různé modely. V mnohých případech mohou být pro účely studie stability komplexní závislosti typu  $x(t) = F(T, t)$  rozloženy do vztahu ve formátu

$$x(t) = F(T, t) = f(T) g(t)$$

**PŘÍKLAD** Obrázek 7 znázorňuje experimentálně získanou závislost organického sumačního parametru u environmentálního CRM.

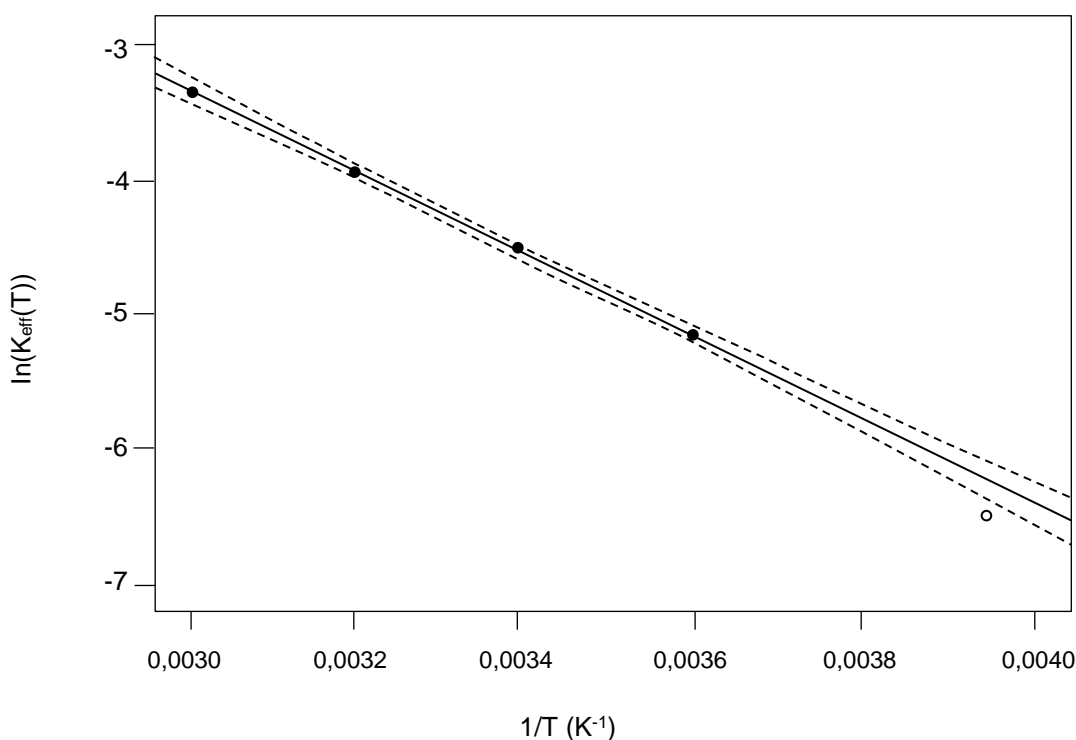
planned storage conditions may be assumed to be negligible subject to confirmation by appropriate monitoring arrangements (see 8.10).

Extrapolation beyond the range of storage conditions tested (for example, predicting degradation rates at  $-20\text{ °C}$  from an experiment involving only temperatures above  $0\text{ °C}$ ) can be unreliable and is not recommended.

**NOTE 1** Processes behind degradation/deterioration of a material can be complex and may require different models. In many cases, complex dependencies of the form  $x(t) = F(T, t)$  can, for stability study purposes, be decomposed into a relationship of the form

$$x(t) = F(T, t) = f(T) g(t)$$

**EXAMPLE** Figure 7 shows an experimentally obtained dependency for an organic sum parameter in an environmental CRM.



**POZNÁMKA** Obrázek znázorňuje Arrheniovu závislost u RM certifikovaného na celkový obsah ropných uhlovodíků (TPH) v půdě. Obrázek ukazuje závislost efektivní degradační rychlosti  $k_{eff}(T)$  (plné body ●) na převrácené hodnotě teploty a 95% konfidenční interval přímky. Uvádíme, že finální bod (prázdný kroužek o) nebyl z technických důvodů vyloučen z regrese, nicméně regresní přímka je prodloužena, aby ukázala předpovídanou změnu při této teplotě a lépe ilustrovala konfidenční meze. Fitovaný vztah vyhovuje dobře při vyšších teplotách představovanými prvými čtyřmi body.

**NOTE** The figure shows an Arrhenius plot for an RM certified for total petroleum hydrocarbon (TPH) content in soil. The figure shows the dependence of the effective degradation rate  $k_{eff}(T)$  (solid circles ●) on the inverse temperature and the 95 % confidence interval for the line. Note that the final data point (empty circle o) was excluded from the regression for technical reasons; the regression line is nonetheless extended to show the predicted change at that temperature and better illustrate the confidence limits. The fitted relationship is good for the higher temperatures represented by the first four points.

**Figure 7 — Results for an accelerated stability study**

**Obrázek 7 — Výsledky urychlené studie stability**



POZNÁMKA 2 Další příklady aplikací uvádějí a zevrubně popisují odkazy <sup>[23]</sup> a <sup>[24]</sup>.

### 8.5.5 Další zdroje náhodné variability při studiích stability

Často se uplatňuje více než jeden náhodný vliv, například pokud jsou prováděna pozorování u mnoha jednotek RM duplicitně při každém časovém bodu, pak (v principu) existuje člen mezi jednotkami a rovněž člen reziduální chyby. Vyžaduje se pak speciální zpracování za použití statistického softwaru k tomu, aby se zcela zahrnula taková struktura chyb. Nejběžnějším přístupem je modelování smíšených vlivů pomocí metod maximální věrohodnosti, které počítají s násobnými náhodnými vlivy a vypočítávají rozptyly těchto náhodných vlivů a požadované směrodatné chyby stálých vlivů (směrnice a úsek) podle.<sup>[25],[26]</sup> Avšak jednoduché zpracování, je-li počet replikátů identický u jednotky RM, je zprůměrovat násobné replikáty tak, aby poskytly jednu hodnotu pro každou jednotku RM; toto může být zpracováno jako statistický model s jedním reziduálním chybovým členem. Nevýhodou je, že redukuje zdánlivý počet stupňů volnosti a může tudíž zvyšovat nejistotu přidruženou koeficientům regrese.

### 8.6 Opatření při zjištění významného trendu ve studii stability

Pokud je pozorován významný trend (viz 8.5.2.6), měl by se uplatnit jeden z následujících přístupů:

- hodnota vlastnosti, u které byla pozorována degradace, nebude certifikována,
- období platnosti certifikované hodnoty bude zkráceno na základě předpovědi modelu, aby se zajistilo, že změna hodnoty nebude technicky významná,
- odhadne se očekávaný rozsah degradace v zamýšlené době platnosti, převede se na standardní nejistotu a zahrne se spolu s nejistotou očekávané degradace do nejistoty přidělené hodnoty,
- certifikovaná hodnota a její nejistota se uvedou jako funkce času a odrazí tak odhadovaný trend a jeho nejistotu,
- kombinaci dvou či více přístupů podle b), c) a d).

POZNÁMKA 1 Přístup podle c) je rovnocenný zahrnutí nekorigovaného vychýlení (bias) do nejistoty měření.

POZNÁMKA 2 Přístup podle d) vyžaduje dobře stanovit model degradace.

POZNÁMKA 3 Odstavec 8.7.4 poskytuje další podrobnosti jak vyhodnocovat nejistotu, je-li přítomen významný trend.

PŘÍKLAD 1 CRM je certifikován na molární koncentraci <sup>242</sup>Pu v roztoku. Tato koncentrace se snižuje s časem v

NOTE 2 More application examples are given, and comprehensively described, in References <sup>[23]</sup> and <sup>[24]</sup>.

### 8.5.5 Additional sources of random variation in stability studies

Often, there may be more than one random effect; for example if multiple RM units are observed in duplicate at each point in time, there is (in principle) a between-unit term, as well as the residual error term. Specialised treatment, ideally using statistical software, is required to take full account of such an error structure. The most common approach is mixed-effects modelling using maximum likelihood methods, which allows for multiple random effects and calculates the variances for the random effects and the required standard errors for fixed effects (slope and intercept), accordingly. <sup>[25],[26]</sup> However, a simple treatment when the number of replicates per RM unit is identical is to average multiple replicates to give a single value per RM unit; this can then be treated as a statistical model with a single residual error term. The disadvantage is that this reduces the apparent number of degrees of freedom and may consequently increase the uncertainty associated with regression coefficients.

### 8.6 Action on finding of a significant trend in a stability study

If a technically significant trend (see 8.5.2.6) is observed, one of the following approaches should be adopted:

- the property value for which the degradation was observed is not certified;
- the period of validity of the certified value is decreased, based on the model prediction, to ensure that the change in value is not technically significant;
- the expected extent of degradation over the intended period of validity is estimated, converted into a standard uncertainty and included, together with the uncertainty of the expected degradation, in the uncertainty of the assigned value;
- the certified value and its uncertainty are given as a function of time, reflecting the estimated trend and its uncertainty;
- a combination of two or more of b), c) and d).

NOTE 1 The approach in c) above is equivalent to including an uncorrected bias in a measurement uncertainty.

NOTE 2 Approach d) above requires a well-established degradation model.

NOTE 3 Subclause 8.7.4 gives further details of the evaluation of uncertainties in the presence of a significant trend.

EXAMPLE 1 A CRM is certified for the molar concentration of <sup>242</sup>Pu in solution. This concentration decreases over time

důsledku dobře známého radioaktivního rozpadu tohoto izotopu. V certifikátu se udává, že hodnota je platná pro určitý den (den, ve kterém byla provedena měření) a udává roční ztrátu izotopu (0,000 22 % za rok), spolu s použitým poločasem rozpadu a nejistotou tohoto poločasu, aby byl možný výpočet přesné hodnoty k datu použití a rovněž nejistoty této hodnoty platné v té době.

**PŘÍKLAD 2** Zjistilo se, že u RM pro kvantitativní mikrobiologii dochází v průměru ke ztrátě 0,05 log KTJ (cfu) za měsíc s nejistotou této ztráty 0,01 log KTJ (cfu) za měsíc (jako standardní nejistota), a ztráta je považována za významnou ve srovnání s hodnotou vlastnosti. Počáteční hodnota vlastnosti je hodnota získaná v době charakterizace. Aby se započítala ztráta při skladování, je vydáván certifikát s korigovanou hodnotou na dobu vydání a nejistota korekce je zahrnuta v deklarované nejistotě hodnoty vlastnosti (viz 8.7.4).

## 8.7 Vyhodnocení nejistoty ze studie stability

### 8.7.1 Obecné úvahy pro vyhodnocení nejistoty ze studií stability

Tam, kde platné technické důvody prokazují, že potenciální změna v průběhu doby platnosti certifikátu je zanedbatelná ve srovnání s certifikovanou nejistotou a to opírá o zkušenosti a pozorování, pak složka nejistoty patřící dlouhodobé stabilitě může být nastavena na nulu nebo vynechána z nejistoty certifikované hodnoty.

Za jiných okolností, kdy analýza dat o stabilitě poskytuje odhadovanou rychlost změn, obvykle odvozenou jako nebo z koeficientů fitovaného modelu, pak je možné předpovědět potenciální budoucí změnu na základě takového modelu. Tyto předpovědi mohou formovat výběr režimu monitoringu. Kromě toho může být možné provést odhad nejistoty spojené s předpovídanými změnami v certifikovaných hodnotách v průběhu času. Je-li k dispozici jen málo předchozích informací o chování materiálu po delší dobu (tedy delších než je doba trvání studie stability) a výrobce zvolí použít relativně málo časté monitorování (například ročně nebo méně), je rozumné odhadnout tyto nejistoty a u certifikovaných referenčních materiálů je začlenit do nejistoty přidružené certifikované hodnotě (hodnotám).

Tento odstavec (8.7) poskytuje základní návod pro zdroje nejistot, které by se měly zvažovat za těchto okolností a obecný postup pro odhad nejistoty spojené s možnou změnou s časem.

**POZNÁMKA 1** Degradace matrice může ovlivňovat měřené hodnoty získané některými zkušebními metodami i tam, kde (pravá) hodnota vlastnosti zůstane stabilní.

**POZNÁMKA 2** Odhad stability na základě platného technického zdůvodnění musí zvážit velikost možné změny ve vztahu k velikosti nejistoty vyžadované pro zamýšlené použití materiálů.

due to the well-established radioactive decay of the isotope. The certificate states that the value is valid for one exact day (the day on which the measurements were performed) and states the yearly loss of the isotope (0,000 22 % per year), with the half-life used and the uncertainty of the half-life, to allow calculation of an exact value on the date of use as well as an uncertainty for the value at that time.

**EXAMPLE 2** An RM for quantitative microbiology is found to lose, on average, 0,05 log cfu per month of storage, with an uncertainty of this loss of 0,01 log cfu per month (as standard uncertainty), and the loss is considered important compared to the uncertainty of the property value. The initial property value is the value obtained at the time of characterization. To allow for loss on storage, certificates are issued with a corrected value valid at the time of issue and the uncertainty of the correction is included in the claimed uncertainty for the property value (see 8.7.4).

## 8.7 Uncertainty evaluation from stability studies

### 8.7.1 General considerations for uncertainty evaluation from stability studies

Where valid technical reasons demonstrate that the potential change over the period of validity of the certificate is negligible compared to the certified uncertainty, and this is supported by experience and observation, then the component of uncertainty due to long-term stability may be set to zero or omitted from the uncertainty in the certified value.

In other circumstances, where stability data analysis produces estimated rates of change, usually derived as, or from, coefficients in a fitted model, it becomes possible to predict potential future change based on the model. Such predictions can inform the choice of monitoring regime. In addition, it becomes possible to make an estimate of the uncertainty associated with predicted changes in certified values over time. Where little prior information is available about the behaviour of the material over extended periods of time (that is, larger than the stability study duration), and the producer chooses to employ comparatively infrequent monitoring (e.g. yearly or less), it is prudent to estimate these uncertainties and, for certified reference materials, to include them in the uncertainty associated with the certified value(s).

This subclause (8.7) provides basic guidance on the sources of uncertainty to be considered in these circumstances and on the general procedure for estimating uncertainty associated with possible change over time.

**NOTE 1** Matrix degradation can affect measured values obtained by some test methods even where the (true) value of the property remains stable.

**NOTE 2** Judgement about stability based on valid technical reasons, supported by experience and observation, needs to consider the size of possible changes in relation to the size of the uncertainty required for the intended use of the material.

### 8.7.2 Zdroje nejistoty předpovídané změny v čase

Studie stability počítá s následujícími zdroji variability:

- a) opakovatelností měření,
- b) variabilitou mezi sériemi u měřicího systému,
- c) heterogenitou mezi jednotkami (při charakterizaci šarže).

Všechny zdroje variability (a až c, výše) přispívají k nejistotě při předpovědích s výjimkou, že při izochronních návrzích vystupuje variabilita mezi sériemi jen v tom případě, když jsou nutností vícenásobné měřicí série. Tam, kde se tyto zdroje variability uplatňují a přispívají k variabilitě výsledků měření během studie, směrodatné chyby odhadovaných koeficientů modelu budou, pokud jsou vhodně fitovány, zohledňovat tyto nejistoty.

Nejistoty přidružené náhodné variabilitě během studie se musí zahrnout do každého odhadu nejistoty přidružené k předpovídané změně v čase.

Tam, kde předpovídaná změna používá několik modelových koeficientů, odhadovaná nejistota by měla počítat s kompletní kovarianční maticí použitých koeficientů, protože chyby v modelu jsou často silně korelovány.

Kromě toho jsou přítomny systematické vlivy včetně (ale neomezuje se na tyto):

- nejistoty v modelových koeficientech vyplývající z měření času nebo měření odezvy,
- nejistoty plynoucí z výběru modelu, například předpoklad lineární degradace spíše než exponenciální změny nebo (přísněji) namísto autokatalytického modelu.

Příspěvky patřící systematickým vlivům na měření času a odezvy jsou obvykle mnohem menší než ty pocházející z náhodné variability. Není běžnou praxí zahrnovat tyto příspěvky do odhadu nejistot předpovědi používaných pouze k vyhodnocení nejistoty, i když je prozíravé to tak činit, když se jedná o odhad korekce, která bude součástí certifikátu nebo návodu k použití.

Model používaný pro extrapolace z podmínek urychlené studie na vlivy v reálném čase při definovaných podmínkách skladování musí být platný a musí se zvažovat nejistota tohoto modelu. Nejistoty spojené s výběrem výchozího modelu se obvykle nezačleňují do odhadu nejistot predikce. Namísto toho se fitují alternativní kandidátské modely a využívají se jejich předpovědi pro volbu režimu monitorování.

POZNÁMKA Odstavec 8.5.5 se zabývá zpracováním násobných zdrojů variability u studií stability.

### 8.7.2 Sources of uncertainty in predicted change over time

A stability study includes the following sources of variation:

- a) repeatability of measurement;
- b) between-run variability of the measurement system;
- c) between-unit heterogeneity (in batch characterization).

All the sources of variation (a to c, above) contribute to uncertainty in prediction, with the exception that in isochronous designs run-to-run variability is only present where multiple measurement runs are necessary. Where these sources of variation are present and contribute to the variation in measurement results during the study, the standard errors in estimated model coefficients will, if properly fitted, take due account of these uncertainties.

Uncertainties associated with random variability during the study shall be included in any estimate of the uncertainty associated with a predicted change over time.

Where a predicted change uses several model coefficients, the estimated uncertainty should take account of the complete covariance matrix for the coefficients used, as errors in model coefficients are often highly correlated.

In addition, systematic effects are present, including (but not limited to):

- uncertainties in model coefficients arising from measurement of time, or measurement of response;
- uncertainty arising from the choice of model, for example an assumption of linear degradation rather than exponential change or (more severely) instead of an autocatalytic model.

Contributions associated with systematic effects on the measurement of time and response are usually much smaller than those arising from random variation. It is not common practice to include these contributions in estimating prediction uncertainties used solely for uncertainty evaluation, though it is prudent to do so when estimating a correction to be included on a certificate or in instructions for use.

The model used for extrapolation from the conditions of the accelerated study to the effects in real time under defined storage conditions, has to be valid and the uncertainty of this model has to be considered. Uncertainties associated with choice of underlying model are not usually included in estimating prediction uncertainties. Instead, alternative candidate models may be fitted and their predictions used to inform the selection of the monitoring regime.

NOTE Subclause 8.5.5 discusses the treatment of multiple sources of variation in stability studies.

### 8.7.3 Odhad nejistot stability za nepřítomnosti významných trendů

Tam, kde existuje technické zdůvodnění stability, je možné předpokládat, že materiál je stabilní a že nejistota spojená se stabilitou může být nastavena na nulu spolu s přijetím režimu monitoringu, který může rychle detekovat neočekávané změny, ideálně dříve než mohou mít nepříznivý vliv na použití materiálu. Volbou počátečního bodu monitoringu a jeho následnými intervaly se zabývá odstavec 8.10.

Tam, kde chybí technické zdůvodnění stability, které by bylo podpořeno zkušenostmi a kde výrobce zvolí použití delších intervalů monitoringu, měla by se odhadnout nejistota spojená s možnou nestabilitou a začlenit do nejistoty přidružené ke každé certifikované hodnotě. Taková nejistota spojená s možnou nestabilitou by pak měla vycházet z nejistoty spojené s předpovědí změny hodnoty v čase rovném prvnímu bodu monitoringu plus době expirace jakéhokoliv certifikátu vydaného do té doby. Předpověď by měla používat vhodný model fitovaný pomocí dat o stabilitě, které jsou k dispozici a nejistota by měla brát v úvahu všechny odpovídající zdroje náhodné variability (viz 8.7.2).

U jednoduchého lineárního modelu používaného u klasických studií stability s několika časovými body, je nejistota  $u_{\text{Its}}$  přidružená předpovědané změně dána jako

$$u_{\text{Its}} = s(b_1)(t_{m1} + t_{\text{cert}}) \quad (11)$$

kde  $s(b_1)$  je směrodatná chyba odhadované směrnice, vypočtená jako v dodatku B.3,  $t_{m1}$  je časový interval mezi přidělením hodnoty a počátečním bodem monitorování stability a  $t_{\text{cert}}$  je životnost certifikátu vydaného v té době.

Někteří výrobci neuvádějí datum expirace vztažené ke každému vydanému certifikátu a dávají přednost stanovení data expirace vycházejícího z plánované životnosti materiálu. V těchto případech člen  $(t_{m1} + t_{\text{cert}})$  ve vzorci (11) může představovat plánovanou životnost (to je  $t_{\text{cert}}$ ) nebo podle zvážení výrobce dobu odpovídající bodu druhého plánovaného monitoringu s tím, že se spoléhá na to, že uživatelé budou neprodleně informováni o jakékoliv změně, která se zjistí po počátečním bodu monitoringu.

POZNÁMKA Zdůvodnění vzorce (11) se uvádí v odkazu <sup>[18]</sup>.

### 8.7.4 Odhad nejistot stability v případech známého významného trendu

Tam, kde existuje známý statisticky významný nebo technicky významný trend, výrobce může podle odstavce 8.6 poskytnout časově závislou certifikovanou hodnotu a zvýšit nejistotu predikce. V tomto odstavci se uvádějí podrobnosti o vyhodnocení nejistoty pro tyto dva případy.

### 8.7.3 Estimation of stability uncertainties in the absence of significant trends

When there is technical justification for stability, it may be assumed that the material is stable and the uncertainty associated with stability may be set to zero along with adopting a monitoring regime that can detect unexpected change promptly, ideally before it can adversely affect the use of the material. The choice of initial monitoring point and subsequent intervals are discussed in 8.10.

Where there is no technical justification for stability that is supported by experience, and where the producer elects to use longer monitoring intervals, an uncertainty associated with possible instability should be estimated and included in the uncertainty associated with any certified value. The uncertainty associated with possible instability should then be based on the uncertainty associated with prediction of the change in value at a time equal to the first monitoring point plus the time to expiry of any certificate issued up to that time. The prediction should use an appropriate model fitted to the available stability data and the uncertainty should take due account of all relevant sources of random variation (see 8.7.2).

For a simple linear model applied to a classical stability study over several points in time, the uncertainty  $u_{\text{Its}}$  associated with the predicted change is given by

$$u_{\text{Its}} = s(b_1)(t_{m1} + t_{\text{cert}}) \quad (11)$$

where  $s(b_1)$  is the standard error for the estimated slope, calculated as in Clause B.3,  $t_{m1}$  is the time interval between value assignment and the initial stability monitoring point and  $t_{\text{cert}}$  is the lifetime of a certificate issued during that time.

Some producers do not provide an expiry date specific to each certificate issued, preferring to set an expiry date based on a planned lifetime for the material. In such cases, the term  $(t_{m1} + t_{\text{cert}})$  in Formula (11) may be set to the planned lifetime (that is,  $t_{\text{cert}}$ ) or, at the producer's discretion, to the time to the second planned monitoring point, relying on immediately informing users of any change discovered after the initial monitoring point.

NOTE The justification for Formula (11) is given in Reference <sup>[18]</sup>.

### 8.7.4 Estimation of stability uncertainties in the case of a known significant trend

Where there is a known statistically significant or technically significant trend the producer may, following 8.6, provide a time-dependent certified value or may provide a time independent certified value and increase the prediction uncertainty. This clause provides further detail on the evaluation of uncertainty in these two cases.

Když se poskytuje časově závislá hodnota, funkce pro certifikovanou hodnotu by měla odrážet nejlepší odhad trendu a funkce pro nejistotu by měla reflektovat nejistotu trendu (nebo podle okolností nejistotu korekce) s tím, že se berou v úvahu zdroje nejistoty vyjmenované v odstavci 8.7.2.

Je-li certifikovaná hodnota udávána bez závislosti na čase, měl by výrobce zvýšit nejistotu predikce a umožnit tak zohlednit očekávanou změnu. Zvolí se období platnosti (obvykle  $(t_{m1} + t_{cert})$  jako v 8.7.3) a odhadne se rozsah degradace za toto dobu. Toto se převede na standardní nejistotu (např. za použití pravouhlého rozdělení, je-li pozorován lineární trend) a tato nejistota se sloučí (podle pravidel šíření nejistoty) s nejistotou předpovídané změny.

**POZNÁMKA** Výsledná nejistota u certifikované hodnoty bez závislosti na čase se může použít jen pro jednu stranu intervalu, protože degradace směřuje pouze jedním směrem a poskytuje tak nesymetrické nejistoty.

### 8.8 Odhad životnosti při skladování (skladovatelnosti) ze studie stability

Tam, kde je možné stanovit u certifikované hodnoty přijatelný rozsah změny v důsledku nedostatku stability, je možné odhadnout životnost při skladování, po kterou se očekává, že hodnota zůstane přijatelná pro použití. Zásady popisuje dodatek B.4.

### 8.9 Návod k použití ve vztahu k managementu stability

Návod k použití materiálu by měl obsahovat podrobné informace o tom, jak by materiál měl být skladován u uživatele, aby se zajistilo, že materiál se nebude během doby platnosti kazit nad deklarovanou nejistotu. Tyto informace by měly zahrnovat doporučené podmínky skladování před otevřením a (pokud výslovně je možno použít materiál více než jednou) po otevření. Měly by být obsaženy i jakékoliv nutné restrikce doby použití po otevření jednotky. Návod k použití může obsahovat instrukce pro ověření integrity referenčního materiálu před prvním použitím, takové ověření může například vycházet z kontroly plomby balení nebo pomocí zařízení na záznam teploty je-li součástí balení.

Tam, kde je dovoleno opakované odebírání podvzorků, měl by návod k použití obsahovat jakákoliv opatření na zabránění kontaminace a opatření pro skladování již otevřených jednotek RM, která by zajistila, že zbylý materiál zůstane vhodný pro další použití a u certifikovaných referenčních materiálů, že deklarovaná nejistota nebude ohrožena. Jestliže hodnoty vlastností mohou být ovlivněny opakovaným odběrem vzorků, například odpařováním, nebo opakovaným zmrazováním, mělo by se to poznamenat v certifikátu.

**POZNÁMKA** Je dobrou praxí poučít koncového uživatele, aby skladoval materiály s vyloučením světla, mimo zdroje tepla a v suchu. Takové instrukce též minimalizují rozsah

When a time-dependent value is provided, the function for the certified value should reflect the best estimate of the trend and the function for the uncertainty should reflect the uncertainty of the trend (or, if appropriate, a correction), taking into account the sources of uncertainty listed in 8.7.2.

When a certified value is given independently of time, the producer should increase the prediction uncertainty to allow for expected change. A period of validity, (usually  $(t_{m1} + t_{cert})$  as in 8.7.3) is chosen and the extent of degradation over that time is estimated. This is converted into a standard uncertainty (e.g. using a rectangular distribution if one observes a linear trend) and this uncertainty is combined (using uncertainty propagation rules) with the uncertainty of the predicted change.

**NOTE** The resulting uncertainty for a time-independent certified value can be applied only to one side of the interval, as degradation tends to be in one direction only, thus resulting in asymmetric uncertainties.

### 8.8 Estimation of storage lifetime (“shelf life”) from a stability study

Where it is possible to set an acceptable level of change due to lack of stability for a certified value, it is possible to estimate a storage lifetime within which the value is expected to remain acceptable for use. The principles are described in Clause B.4.

### 8.9 Instructions for use related to management of stability

The instructions for use of the material should include detailed information on how the material should be stored by the user to ensure that the material does not deteriorate beyond the stated uncertainty during the period of validity. This information should include recommended storage conditions prior to and (explicitly, if the material may be used more than once) after opening. Any necessary restrictions on time of use after opening a unit should be included. The instructions for use may include instructions for verification of the integrity of a reference material prior to first use; such verification might for example, be based on package seal inspection or temperature logging equipment included in the packaging.

Where repeated sub-sampling is permitted, the instructions for use should include any precautions for prevention of contamination and for storage of opened units of the RM that are necessary to ensure that the remaining material remains fit for use and, for CRMs, that the stated uncertainty is not compromised. If property values can be affected by repeated subsampling, for example, by evaporation or by repeated refreezing, this should be noted on the certificate.

**NOTE** Instructing the end user to store reference materials away from light, away from sources of heat and in dry conditions is good practice. Such instructions also minimize

podmínek, které vyžadují experimentální studii k ověření požadovaných skladovacích podmínek.

the range of conditions that require experimental study to verify required storage conditions.

## 8.10 Monitorování stability

## 8.10 Stability monitoring

### 8.10.1 Požadavky na monitorování stability

S monitorováním by se mělo počítat po dobu životnosti daného RM. Jak se uvádí v odstavci 8.1, chování daného RM po jeho dobu životnosti se z typických studií stability dá těžko předpovědět. Protože je těžko předpovědět chování, je obvykle nutno monitorovat stabilitu materiálu. Monitorování materiálu po jeho uvedení na trh je tedy důležitou součástí celkového managementu stability referenčních materiálů s dlouhou životností.

### 8.10.1 Requirements for monitoring

Monitoring should be planned for during the lifetime of the RM. As noted in clause 8.1, the behaviour of a given RM over its lifetime is difficult to predict reliably from typical stability studies. Because the behaviour is hard to predict, it is usually necessary to monitor the stability of the material. Monitoring of a material following release is accordingly an important part of the overall management of stability for reference materials with long usable life.

Pokud mají být využity zkušenosti se stabilitou předchozích vyráběných šarží RM k formování tvrzení o stabilitě pro budoucí šarže včetně skladovatelnosti, rychlosti degradace nebo nejistoty způsobené dlouhodobým skladováním, je podstatné mít důkazy podporující taková tvrzení. V těchto případech monitorovací zkoušky by měly zahrnovat měření prováděná v datech expirace předchozích šarží.

If experience of stability from previous RM production batches is to be used to inform stability claims for future batches, including shelf life, rates of degradation, or uncertainty due to long-term storage, it is essential to have evidence to support claims of stability. In these cases, the monitoring tests should include a measurement made at the expiry date for previous batches.

Monitorování není vždy nezbytné. Některé materiály jsou certifikovány pro použití ve velmi krátké životnosti a s krátkou expirací, s náhradou materiály téměř kontinuálně vyráběnými. Takové materiály expirují dříve, než dosáhnou jakýkoliv rozumný bod pro monitorování. Výrobce by měl tedy posoudit potřebu monitorování po uvolnění materiálu. Takové posouzení by mělo zvážit:

Monitoring is, however, not always essential. Some materials are certified for use with very short lifetime and early expiry dates, with replacement materials in near continuous production. Such materials expire before any reasonable monitoring point. The producer should therefore assess the need for post-release monitoring. The assessment should consider:

- trvání a výsledky studií stability,
- zkušenosti se stabilitou příbuzných materiálů (nebo předchozích šarží v podstatě identického materiálu,
- očekávanou prodejní životnost nebo dostupnost daného materiálu,
- zdali prohlášení o nejistotě předmětných vlastností obsahuje toleranci změny,
- sílu důkazu pro předpoklad, že hodnota (hodnoty) vlastností zůstanou stabilní.

- the duration and results of stability studies;
- experience of the stability of related materials (or previous batches of essentially identical material);
- the expected sales lifetime or availability lifetime of the material;
- whether allowance for change is to be included in any statement of uncertainty for the properties of interest;
- the strength of evidence for the assumption that the property value(s) will remain stable.

Monitorování není nutné tam, kde očekávaná životnost je krátká ve srovnání se známými rychlostmi degradace stejných nebo velmi podobných materiálů. U většiny dalších případů je monitorování běžně považováno za potřebné a mělo by se podnikat přinejmenším jednou během životnosti daného materiálu pro potvrzení stability.

Monitoring is not necessary where the expected lifetime is short compared to known degradation rates for the same or closely similar materials. In most other cases, some monitoring is normally considered necessary and should be undertaken at least once over the lifetime of the material to confirm stability.

### 8.10.2 Výběr počátečního bodu monitorování a intervalu (intervalů) monitorování.

### 8.10.2 Choice of initial monitoring point and monitoring interval(s)

#### 8.10.2.1 Důležitost předchozích informací

#### 8.10.2.1 Relevance of prior information

Tam, kde monitorování je považováno za nezbytné, výrobce by měl stanovit počáteční bod pro monitorování a intervaly pro jeho další provádění. Obě volby závisejí ve značné míře na dostupnosti (nebo

Where monitoring is considered necessary, the producer should set an appropriate initial monitoring point and intervals for further monitoring. Both choices depend heavily on the availability (or otherwise) of

jinak) informací po dlouhou dobu získávaných o velmi podobných materiálech.

#### **8.10.2.2 Plány monitorování jsou-li k dispozici obsáhlé předchozí informace**

Tam, kde existují platné relevantní informace z testování stability a/nebo monitorování velmi podobných materiálů po dobu podobnou (nebo delší než) očekávaná životnost dotyčného materiálu, může výrobce stanovit počáteční bod pro monitorování a intervaly, které jsou

- podobné těm, které se úspěšně používají u velmi podobných materiálů, nebo
- založeny na pozorované změně s časem u předchozích materiálů.

Ve druhém případě by se měly intervaly stanovit tak, aby důvodně očekávaná změna mezi body monitorování, vycházející z předchozích informací, nebyla více než třetinou nejistoty přidružené k certifikovaným hodnotám.

Při hodnocení možných změn by se mělo brát v úvahu každé experimentální hodnocení, například výběr kratších intervalů, je-li nejistota veliká. Při posuzování možné změny s časem a pro volbu intervalu monitorování se mohou brát v úvahu relevantní předchozí informace z publikované literatury, zkušenosti, nebo další zdroje.

#### **8.10.2.3 Plány monitorování nejsou-li k dispozici obsáhlé předchozí informace**

##### **8.10.2.3.1 Obecné úvahy o monitorování nejsou-li k dispozici předchozí informace**

Tam, kde není dostatek informací z testování stability a/nebo monitorování blízce podobných materiálů a zvláště když žádný podobný materiál není k dispozici po dobu trvání studie stability dotyčných materiálů, výrobce by měl stanovit počáteční body pro monitorování vycházející z výsledků studie stability dotyčného materiálu a kromě toho plán relativně častého monitorování pro přinejmenším prvé tři monitorovací body.

Existují dvě základní strategie pro výběr počátečních bodů monitorování, když neexistují důkladné údaje o blízce podobných materiálech:

- a) předpovědět možnou změnu a stanovit počáteční bod monitorování dříve než jakákoliv změna nepříznivě ovlivní koncové použití,
- b) použít jednoduchý násobek trvání studie stability.

V odstavcích 8.10.2.3.2 a 8.10.2.3.3 se uvádějí další podrobnosti o těchto strategiích.

V obou případech a) i b) intervaly pro následující body monitorování (tedy po prvních třech bodech) by se

information acquired over extended periods on closely similar materials.

#### **8.10.2.2 Monitoring plans where extensive prior information is available**

Where there is sound, relevant information from stability testing and/or monitoring of closely similar materials, covering a period similar to (or longer than) the expected lifetime of the material in question, the producer may set an initial monitoring point and intervals that are

- similar to those used successfully on closely related materials, or
- based on the observed change over time for the previous materials.

In the latter case, intervals should be set so that reasonably expected change between monitoring points, based on prior information, is not more than one third of the uncertainty associated with certified values.

In assessing the possible change, the uncertainty of any experimental assessment should be taken into account, for example, by selecting shorter intervals when the uncertainty is large. Relevant prior information from published literature, experience or other sources may also be taken into account in assessing the possible change over time and for selecting a monitoring interval.

#### **8.10.2.3 Monitoring plans where extensive prior information is not available**

##### **8.10.2.3.1 General considerations for monitoring where prior information is not available**

Where there is insufficient information from stability testing and/or monitoring of closely similar materials, and in particular, where no similar material has been available for longer than the duration of stability studies carried out on the material(s) in question, the producer should set initial monitoring points based on the stability study results for the material in question, and in addition plan for comparatively frequent monitoring at least for the first three monitoring points.

There are two basic strategies for choosing initial monitoring points in the absence of extended data on closely similar materials:

- a) Predict possible change and set the initial monitoring point prior to any change that adversely affects end use;
- b) Use a simple multiple of the stability study duration.

Subclauses 8.10.2.3.2 and 8.10.2.3.3 provide further details of these strategies.

In both cases a) and b), intervals for subsequent monitoring points (that is, after the first three points)

měly stanovit po provedené revizi výsledků měření u prvních třech bodů monitorování.

POZNÁMKA Strategie b) není vhodná, jestliže trvání dlouhodobé studie stability,  $t_{ITS}$ , je dlouhé ve srovnání s očekávanou životností daného materiálu.

### 8.10.2.3.2 Použití předpovědi změny ke stanovení počátečního bodu monitorování

Pro použití předpovědi změny ke stanovení počátečního bodu monitorování a intervalu [strategie a) v 8.10.2.3.1] se vyžaduje specifikovat toleranci certifikované hodnoty (hodnot). Postup je v principu jasný. Nejprve se zkonstruuje 95% dvoustranný interval pro změnu certifikované hodnoty pro sérii období po přidělení hodnoty. Zadruhé se určí nejbližší bod, při kterém jedna z těchto mezí protíná meze specifikované tolerance, buď graficky, nebo numericky. Tento bod, nebo příhodný dřívější čas se vezme jako první bod monitorování. Podrobnosti uvádí dodatek B.4

POZNÁMKA Specifikovaná tolerance pro referenční materiál může vycházet ze zvážení zamýšleného použití. U CRM takový interval vychází obvykle z rozšířené nejistoty. Například volba tolerance rovnající se jedné třetině rozšířené nejistoty představuje nízké riziko, že se certifikovaná hodnota dostane mimo rozšířenou nejistotu před prvním bodem monitorování.

### 8.10.2.3.3 Použití jednoduchého násobku trvání studie stability

Tato strategie [strategie b) v 8.10.2.3.1], je jednoduchá, ale může poskytovat kratší intervaly než strategie a). Je založena na násobcích trvání dlouhodobé studie stability  $t_{ITS}$ .

PŘÍKLAD Příklad užití strategie b) je uveden níže.

- Nastavte první bod monitorování nejpozději na datum přidělení hodnoty plus  $t_{ITS}$  a datum konce dlouhodobé studie stability plus  $t_{ITS}$ .
- Nastavte dva následující body monitorování v intervalu  $2t_{ITS}$  od prvního.

## 8.10.3 Experimentální návrh

### 8.10.3.1 Klasický návrh monitorování

Monitorování se často uskutečňuje za použití klasického experimentálního návrhu. To představuje měření jednotek referenčních materiálů při normálním skladování v plánovaných časových bodech. Vyhodnocení výsledků spočívá v porovnání každého (průměru) výsledku monitorování s certifikovanou hodnotou a tak v čase kontrolovat jakýkoli významný trend v pozorovaných hodnotách.

Výhodou klasického návrhu je jeho jednoduchost, hlavní nevýhodou pak to, že výsledky mohou být ovlivněny dlouhodobou variabilitou měřicího procesu.

should be set following review of the results of measurement at the first three monitoring points.

NOTE Strategy b) is not appropriate if the long-term stability study time,  $t_{ITS}$ , is long compared to the expected lifetime of the material.

### 8.10.2.3.2 Use of a predicted change to set the initial monitoring point

To use a predicted change to set the initial monitoring point and interval [strategy a) in 8.10.2.3.1], a specified tolerance for the certified value(s) is required. The procedure is straightforward in principle. First, a 95 % two-sided confidence interval for the change in certified value is constructed for a series of times following value assignment. Second, the earliest point at which one of these limits intersects the limits of the specified tolerance is determined, either graphically or numerically. This point, or a convenient earlier time, is taken as the first monitoring point. Details are given in Clause B.4.

NOTE The specified tolerance for a reference material may be based on considerations for intended use. For a CRM, such an interval is usually based on the expanded uncertainty. For example, choosing a tolerance of one third of the expanded uncertainty gives a low risk of the certified value moving outside of the expanded uncertainty prior to the first monitoring point.

### 8.10.2.3.3 Use of a simple multiple of the stability study duration

This strategy [strategy b) in 8.10.2.3.1], is simple but may result in shorter intervals than strategy a). It is based on multiples of the long-term stability study duration  $t_{ITS}$ .

EXAMPLE An example of an application of strategy b) is given below.

- Set the first monitoring point at the later of the value assignment date plus  $t_{ITS}$  and the date of the end of the long-term stability study plus  $t_{ITS}$ .
- Set two subsequent monitoring points at intervals of  $2t_{ITS}$  from the first.

## 8.10.3 Experimental

### 8.10.3.1 Classical monitoring design

Monitoring often takes place using the classical design. This involves measurement of RM units in normal storage at planned points in time. The evaluation of the results involves a comparison of each (mean) monitoring result with the certified value and, over time, a check for any significant trend in the observed value.

The advantage of the classical design is simplicity; the principal disadvantage is that the results can be adversely affected by long-term variations in the measurement process.



### 8.10.3.2 Vyhodnocení výsledků monitorování stability

Základní vyhodnocení jednotlivého experimentu monitorování stability u CRM představuje porovnání nové naměřené hodnoty s certifikovanou hodnotou. Tento přístup vyžaduje znalost standardních nejistot  $u_{\text{mon}}$  a  $u_{\text{CRM}}$  přidružených k  $x_{\text{mon}}$  a  $x_{\text{CRM}}$ , v uvedeném pořadí a příslušného koeficientu rozšíření  $k$  na konfidenční úrovni 95 %. Při použití této metody, není-li splněna podmínka

$$|x_{\text{CRM}} - x_{\text{mon}}| \leq k \sqrt{u_{\text{CRM}}^2 + u_{\text{mon}}^2} \quad (12)$$

pak by se měl učinit závěr, že existuje důkaz o nestabilitě.

Tam, kde jsou kromě certifikované hodnoty k dispozici předchozí výsledky monitorování stejné hodnoty, měla by se provést kontrola trendu. Kontrola trendu v hodnotách tří nebo více bodů monitorování (včetně certifikované hodnoty) se může provést jednoduchou lineární regresí. Může se použít vážení, pokud se nejistoty u jednotlivých bodů významně liší. Tam, kde gradient je významný při 95% konfidenční úrovni, mělo by to vést k závěru, že existuje trend hodnot.

Tam, kde kritérium ve vzorci (12) není splněno, nebo kde existuje důkaz o trendu, lze usuzovat na významnou degradaci materiálu a měla by být přijata příslušná opatření. Mezi možné způsoby opatření patří

- provedení ověřovacích studií (včetně nebo bez dočasného zastavení distribuce RM),
- zastavení distribuce a vyřazení materiálu,
- recertifikace materiálu.

Po následném vyhodnocení mohou být k nashromážděným údajům o materiálu přidány nové údaje z monitorování, aby se revidovaly budoucí body monitorování nebo se opravila doba platnosti certifikované hodnoty.

### 8.10.3.3 Izochronní monitorování stability

Alternativou ke klasickému monitorování je použití varianty izochronního návrhu experimentů k provedení druhu semikontinuální studie stability. Ta představuje skladování některých jednotek materiálu za referenčních podmínek, za kterých se předpokládá, že materiál je extrémně stálý, typicky při velmi nízké teplotě. U každého bodu monitorování, jednotky držené za normálních podmínek skladování se měří spolu s jednotkami z referenčních podmínek za podmínek opakovatelnosti (v ideálním případě) měření. Zpracování dat spočívá v následujícím:

- test na významnou změnu ve srovnání s jednotkami skladovanými za referenčních podmínek,
- kontrola a analýza na trendy v rozdílu nebo poměru hodnot mezi vzorky z normálního skladování a vzorky drženými za referenčních podmínek,

### 8.10.3.2 Evaluation of stability monitoring results

The basic evaluation of a single stability monitoring experiment applied to a CRM relies on comparison of the new measured value with the certified value. The approach requires the standard uncertainties  $u_{\text{mon}}$  and  $u_{\text{CRM}}$  associated with  $x_{\text{mon}}$  and  $x_{\text{CRM}}$ , respectively and an appropriate coverage factor  $k$  at a level of confidence of approximately 95 %. Using this method, if the condition

$$|x_{\text{CRM}} - x_{\text{mon}}| \leq k \sqrt{u_{\text{CRM}}^2 + u_{\text{mon}}^2} \quad (12)$$

is not met, then it should be concluded that there is evidence of instability.

Where previous monitoring results on the same value are available in addition to the certified value, a check for a trend in the values should be performed. A check for a trend in the values over three or more monitoring points (including the certified value) may be performed using simple linear regression. Weights may be applied if the uncertainties at different points differ appreciably. Where the gradient is significant at the 95 % level of confidence, it should be concluded that there is evidence of a trend in the values.

Where the criterion in Formulae (12) is not met or where there is evidence of a trend, this indicates significant degradation of the material and action should be taken. Possible modes of action can include:

- performing confirmatory studies (with or without temporary suspension of RM distribution);
- halting distribution and discarding the material;
- re-certification of the material.

Following evaluation, new monitoring data may be added to the collected stability data for the material in order to revise future monitoring points or amend the estimated period of validity of the certified value.

### 8.10.3.3 Isochronous stability monitoring

An alternative to classical monitoring is to use a variant of the isochronous experiment design to carry out a type of semi-continuous stability study. This involves storing some units of the material under reference conditions at which the material is expected to be extremely stable, typically at very low temperatures. At each monitoring point, units kept under normal storage conditions are measured together with units from the reference conditions under (ideally) repeatability conditions of measurement. Data processing should include one or more of the following:

- a test for significant change compared to the units stored under reference conditions;
- inspection and analysis for trends in the difference or ratio in value between samples in normal storage and samples held under reference conditions;

- modelování, které umožňuje zohlednit variabilitu mezi sériemi měřicího procesu.

POZNÁMKA V odkazu [18] se navrhuje alternativní schéma, ve kterém při každém bodu monitorování se nové jednotky RM umístí do referenčních podmínek. Tak se vyhne potřebě skladovat velkou zásobu jednotek RM držných za referenčních podmínek pro možné budoucí požadavky monitorování za cenu komplikovanější analýzy dat.

## 9 Charakterizace materiálů

### 9.1 Úvodní část

Pokyny v této kapitole se týkají především měření zaměřených na přidělení hodnot vlastností materiálu ( $y_{\text{char}}$ ). Studie určující indikativní nebo informační hodnoty (ať tak či onak nazývané), mohou též postupovat podle principů načrtnutých v této kapitole, ale obecně se u nich vyžaduje menší náročnost, zejména s hledem na odhad nejistot měření a zajištění metrologické návaznosti. (viz 9.11).

Certifikované hodnoty vlastností by měly být dobrými odhady pravých hodnot. Certifikované hodnoty mohou být totožné pro mnoho individuálních jednotek (zpracování po šaržích) nebo každé jednotce může být přidělena individuální hodnota v případech, kdy se vyrábí řada jednotlivých artefaktů.

U certifikovaných hodnot spolu s  $y_{\text{char}}$ , se musí stanovit přidružená nejistota charakterizace ( $u_{\text{char}}$ ). Pokyn ISO 34 [2] vyžaduje, aby výrobce poskytl *dokumentovaný důkaz metrologické návaznosti výsledků měření ke stanovené referenci*. To znamená, že ať již se použije jakýkoliv přístup, musí se jasně definovat metrologická návaznost očekávaných hodnot. Toho lze dosáhnout pouze tak, že hodnoty, které se slučují, vskutku určují stejnou měřenou veličinu a výsledky jsou návazné ke stejné metrologické referenci. V ideálním případě jsou takovou referencí Mezinárodní soustava jednotek SI, které je třeba dávat přednost, ale lze použít i jiné reference. Metrologická návaznost se také vztahuje na veličiny definované postupem; zásadní význam pro zajištění návaznosti k definovaným metrologickým referencím má vhodná kalibrace. Odstavec 9.2 poskytuje návod pro zajištění metrologické návaznosti u charakterizace referenčních materiálů.

Charakterizaci lze provádět za použití jedné či více metod v jedné či několika laboratořích. Pokyn ISO 34 [2] vyjmenovává několik základních přístupů k charakterizaci. Se zřetelem na uspořádání studie, návaznost a statistické vyhodnocení, lze tyto přístupy redukovat na následující možnosti:<sup>[27]</sup>

- charakterizace měřené veličiny, která nezávisí na metodě za použití jedné metody v jedné laboratoři,
- charakterizaci měřené veličiny, která nezávisí na metodě za použití dvou či více metod prokazatelně

- modelling that allows for run-to-run variation in the measuring process.

NOTE Reference [18] suggests an alternative scheme in which fresh RM units are placed at the reference temperature at each monitoring point. This avoids the necessity of a large stock of RM units held at the reference temperature for possible future monitoring requirements, at the expense of more intricate data analysis.

## 9 Characterization of the material

### 9.1 Preamble

The guidance in this clause is intended mainly for the measurements performed to assign the certified property values of a material ( $y_{\text{char}}$ ). Studies for the determination of indicative or information values (however named), may also follow the principles outlined in this clause, but will in general require less rigour, especially with respect to estimation of measurement uncertainties and establishment of metrological traceability (see 9.11).

Certified property values should be good estimates of the true values. The certified value may be the same for many individual units (batch processing), or an individual value may be assigned to each unit in cases where a number of single artefacts are being produced.

For certified values, together with  $y_{\text{char}}$ , the associated uncertainty of characterization ( $u_{\text{char}}$ ) shall be determined. ISO Guide 34 [2] requires a reference material producer to provide *documentary evidence on the metrological traceability of the measurement result to a stated reference*. This means that whatever the approach chosen, the metrological traceability of the expected values shall be clearly defined. This can only be achieved if values that are combined really address the same measurand and the results are traceable to the same metrological reference. Ideally, the International System of Units (SI) is the preferred metrological reference, but other references can be used. Metrological traceability also applies to operationally defined quantities; it remains essential to ensure traceability to defined metrological references by proper calibration. Subclause 9.2 gives guidance on establishing metrological traceability for reference material characterization.

Characterization can be achieved by using one or several methods in one or several laboratories. ISO Guide 34 [2] lists several basic approaches to characterization. With respect to the study setup, traceability and statistical evaluation, these approaches can be reduced to the following possibilities:<sup>[27]</sup>

- characterization of a method-independent measurand using a single method in a single laboratory;
- characterization of a method-independent measurand using two or more methods of

přesnosti v jedné či více kompetentních laboratořích,

- charakterizaci měřené veličiny definované postupem za použití skupiny kompetentních laboratoří.

Tato kapitola pojednává o těchto základních principech, rovněž o způsobu provádění mezilaboratorních studií, charakterizaci čistoty a charakterizaci přímým srovnáním s blízce podobnými certifikovanými referenčními materiály. I když současný stav charakterizace kvalitativních a nominálních vlastností (tj. identity látek) není ve stavu, aby se mohly poskytnout podrobné směrnice, některé obecné principy pro takovou charakterizaci jsou též zahrnuty do této kapitoly.

## 9.2 Zajištění metrologické návaznosti

### 9.2.1 Princip

V praxi metrologická návaznost sestává ze dvou částí, a to jasně definice vlastnosti, jejíž hodnoty jsou návazné a návaznosti hodnot vlastností samých k vhodné stanovené referenci. Stanovení návaznosti tedy vyžaduje, aby výrobce zabezpečil identitu měřené vlastnosti, tj. tedy prokázal, že prováděná měření vskutku kvantifikují tuto vlastnost. Výsledky měření se pak musí učinit návaznými ke stanovené referenci použitím řádně zdokumentovaného a validovaného postupu, který poskytuje odhad hodnoty měřené veličiny s přijatelnou přesností.

POZNÁMKA Další pojednání o metrologické návaznosti pro chemická měření uvádí odkaz <sup>[28]</sup>.

### 9.2.2 Metrologické reference

Metrologická návaznost výsledků měření se obvykle zajišťuje vhodnou kalibrací všech relevantních vstupních veličin pomocí příslušných standardů. Kvantitativní vlastnosti mohou být návazné k

- obecně přijímané soustavě jednotek (např. Mezinárodní soustavě jednotek SI) a vyjádřené v odpovídajících jednotkách. To by se mělo uskutečnit všude tam, kde je to možné, nebo
- standardy měření nebo artefakty, včetně certifikovaných referenčních materiálů.

Ve většině případů budou laboratoře používat standardy měření, které budou nositeli hodnot návazných k vyšší referenci (např. SI). Hodnoty získané kalibrací s takovým standardem jsou návazné k této vyšší referenci prostřednictvím standardu, pokud všechny ostatní vstupní faktory byly řádně kalibrovány.

POZNÁMKA 1 Typickou položkou pro kalibraci jsou váhy, teploměry, momentové klíče, volumetrické vybavení a standardy.

demonstrable accuracy in one or more competent laboratories;

- characterization of an operationally defined measurand using a network of competent laboratories

This clause provides guidance on these basic principles as well as on the conduct of collaborative studies, characterization of purity and characterization by direct comparison with closely matched CRMs. While the current state-of-the-art is not sufficiently evolved to give detailed guidelines for the characterization of qualitative and nominal properties (i.e. identity of the substance), some general principles are also listed in this clause.

## 9.2 Establishing metrological traceability

### 9.2.1 Principle

Metrological traceability in practice consists of two parts, namely a clear definition of the property whose values are traceable and the traceability of the property values themselves to the appropriate stated references. Establishing traceability therefore requires a producer to ascertain the identity of the measured property, i.e. demonstrate that the measurements performed indeed quantify this property. The measurement results in turn shall be made traceable to the stated reference by application of a properly documented and validated procedure that provides an estimate of the value of the measurand with acceptable accuracy.

NOTE Further guidance on metrological traceability for chemical measurements is given in Reference <sup>[28]</sup>.

### 9.2.2 Metrological references

The traceability of measurement results is usually ensured through proper calibration of all relevant input quantities against appropriate standards. Quantity values can be traceable to

- a generally accepted system of units [e.g. the International System of Units (SI)] and expressed in the corresponding units. This should be attempted wherever possible, or
- measurement standards or artefacts, including CRMs.

In most cases, laboratories will use measurement standards that carry values traceable to a higher reference (e.g. SI). Values obtained by calibration with such a standard are traceable to this higher reference via the standard, if all other input factors have been duly calibrated.

NOTE 1 Typical items to be calibrated are balances, thermometers, torque-wrenches, volumetric instruments, and standards.

**POZNÁMKA 2** V mnohých případech se používají měřicí standardy a artefakty včetně certifikovaných referenčních materiálů jako kalibrační standardy měřících metod. Mezi příklady patří pracovní standardy (návazné k primárním standardům) a konvenční stupnice jako pH, pro kterou primární realizace je ztělesněna Harnedovým článkem a rutinní kalibrace používají roztoky pufrů. Hodnoty, získané kalibrací s tímto standardem jsou návazné k této vyšší referenci prostřednictvím hodnot daného standardu, pokud všechny ostatní vstupní faktory byly řádně zkalibrovány.

**POZNÁMKA 3** Jestliže CRM použitý ke kalibraci je sám o sobě návazný k vyšší referenci (např. SI) pak nový CRM bude návazný k této vyšší referenci prostřednictvím kalibrace s CRM.

### 9.2.3 Typy měřených veličin

Měřená vlastnost může být

- definována postupem. V tomto případě je měřená veličina definována odkazem na zdokumentovaný a široce přijímaný měřicí postup a jen výsledky získané stejným postupem se mohou porovnávat,
- opakovanou realizací SI jednotky, definované bez odkazu na konkrétní měřicí metodu. Toto je případ základních fyzikálních vlastností (délky, hmotnosti) a koncentrace nepochybně definovaných látek, které vykazují přímou vazbu na mol. V tomto případě má měřená veličina smysl bez odkazu na konkrétní postup měření.

Ať je či není měřená veličina definovaná postupem, stanovení návaznosti vyžaduje stejnou činnost: každou veličinu, která materiálně ovlivňuje výsledky měření, je třeba kalibrovat nebo se musí mít pod vhodnou kontrolou, obvykle používáním kalibrovaných přístrojů.

#### PŘÍKLADY

- a) Hmotnostní zlomek celkového Cd ve vzorku se stanovuje pomocí dvou validovaných metod založených na rozkladu-ICPMS a neutronové aktivační analýze. Výsledky ze dvou měření velmi odlišného principu souhlasí, což poskytuje jistotu, že vychýlení u obou metod je minimální a potvrzuje, že tato měřená veličina je vskutku „hmotnostní zlomek celkového Cd“. Měřená veličina je tedy paralelní realizací základní jednotky SI (molu Cd).
- b) Hmotnostní zlomek vlákniny jak ji definuje ISO 6865 <sup>[29]</sup> se stanovil v mezilaboratorní studii, v níž všichni účastníci používali ISO 6865. Všechny podmínky měření (teploty, objemy, hmotnost atd.) byly patřičně kalibrovány. Měřená veličina je definována postupem (ISO 6865) a výsledky jsou návazné k SI. Není třeba počítat s příspěvkem možného vychýlení (bias).
- c) Hmotnostní zlomek vlákniny se stanovoval blízkou infračervenou spektrometrií (NIR). Přístroj byl kalibrován použitím výsledků měření podle

**NOTE 2** In many cases, measurement standards or artefacts, including CRMs, will be used as calibration standards for the measurement methods. Examples are working standards (traceable to the primary standard) or conventional scales like pH, for which the agreed primary realization is the Harned cell and for which routine calibration uses buffer solutions. Values obtained by calibration with this standard are traceable to this higher reference via the values of the standard, if all other input factors have been duly calibrated.

**NOTE 3** If a CRM used for calibration is itself traceable to a higher reference (e.g. the SI), then the new CRM will be traceable to this higher reference via calibration with the CRM.

### 9.2.3 Types of measurands

A measured property can be

- operationally defined. In this case, the measurand is defined by reference to a documented and widely accepted measurement procedure and only results obtained by the same procedure can be compared;
- a multiple of a realization of an SI unit, defined without reference to a particular procedure for measurement. This is the case for basic physical properties (length, mass) and concentrations of clearly defined substances, which can be directly linked to the mole. In this case, the measurand is meaningful without reference to a particular measurement procedure.

Whether or not the measurand is operationally defined, the establishment of traceability requires the same activity: every quantity that materially affects the measurement result shall be subject to calibration or shall be kept under suitable control, usually by use of calibrated instruments.

#### EXAMPLES

- a) The mass fraction of total Cd in a sample is determined by two validated methods based on digestion-ICPMS and neutron activation analysis. The results from two very different principles of measurement agree, providing confidence that method bias for either procedure is minimal and confirms that the measurand is indeed “mass fraction of total Cd”. The measurand is therefore a multiple of a realization of a base SI Unit (mole of Cd).
- b) The mass fraction of crude fibre as defined by ISO 6865 <sup>[29]</sup> is determined by an interlaboratory study in which all participants apply ISO 6865. All measurement conditions (temperatures, volumes, mass, etc.) were duly calibrated. The measurand is operationally defined (ISO 6865) and results are traceable to the SI. No allowance needs to be made to allow for possible procedure bias.
- c) The mass fraction of crude fibre was determined by near-infrared spectrometry (NIR). The instrument was calibrated using measurement

ISO 6865. Měřená veličina je definována postupem (hmotnostní podíl vlákniny stanovený NIR) a hodnoty veličiny jsou návazné k výsledkům ISO 6865.

#### 9.2.4 Vliv přípravy vzorku nebo předběžné úpravy

U mnoha matricových referenčních materiálů je situace složitá. I když instrumentální stanovení hodnoty vlastnosti lze učinit návazným k příslušným jednotkám pomocí kalibrace měřicího vybavení, kroky předběžné úpravy jakými jsou extrakce, předběžná úprava nebo převod vzorku z jednoho fyzikálního či chemického stavu do druhého se nemůže snadno zkalibrovat. Takové úpravy je možné jen porovnávat s referenčním postupem (pokud existuje), nebo mezi sebou. To činí jasnou definici měřené veličiny poněkud komplikovanou. Obecně existují tři možnosti.

- a) Pro některé úpravy jsou definovány referenční postupy a mohou se použít u certifikačních projektů, aby poskytly certifikovanou hodnotu definovanou odkazem na referenční metodu. To pak poskytuje veličinu definovanou postupem.
- b) Druhou možností je použití dvou či více nezávislých postupů k zhodnocení vychýlení (bias) postupu. Jestliže výsledky z nezávislých postupů navzájem souhlasí v rámci svých příslušných nejistot, může výrobce soudit, že získané hodnoty nejsou významně ovlivněny jednotlivými postupy a tudíž není měřená veličina definována postupem.
- c) V jiných případech je možné pouze porovnání mezi různými laboratořemi, které používají stejný postup. V tomto případě je nemožné prokázat nepřítomnost vychýlení (bias) metody a výsledek je měřená veličina definovaná postupem.

**PŘÍKLAD** Hmotnostní zlomek Cd v půdě byl stanoven v několika laboratořích, které všechny používaly extrakci lučavkou královskou a následnou kvantifikaci ICPMS. Měřená veličina je definována postupem jako ‚získaná extrakcí lučavkou královskou s následnou kvantifikací ICPMS‘.

Definice vlastnosti v poskytované dokumentaci uživatelům musí odrážet zvolený přístup charakterizace.

#### 9.2.5 Ověření návaznosti

V mnoha případech je těžké prokázat patřičnou kalibraci úplně každé součásti vybavení. Takové situace mohou vznikat kvůli neznámým ovlivňujícím faktorům, ale vznikají častěji při charakterizačních studiích s účastí mnoha laboratoří, tam, kde je získání kalibračních certifikátů pro úplně každý použitý přístroj nepraktické. V takových případech adekvátnost opatření k zajištění patřičné kalibrace vybavení a návaznost výsledků by se měly ověřovat například speciálně zařazovanými a připravovanými kontrolními

results obtained following ISO 6865. The measurand is operationally defined (mass fraction of crude fibre as determined by NIR) and the quantity values are traceable to the results of ISO 6865.

#### 9.2.4 Effect of sample preparation or pre-treatment

For many matrix reference materials, the situation is complex. Although the instrumental determination of the property value can be made traceable to appropriate units by the calibration of the measurement equipment used, pre-treatment steps such as extraction, pre-conditioning or transformation of the sample from one physical or chemical state to another cannot easily be calibrated. Such treatments can only be compared with a reference procedure (when available), or among themselves. This makes the clear definition of the measurand somewhat complicated. Generally, three possibilities exist.

- a) For some treatments, reference methods have been defined and may be used in certification projects to provide a certified value defined by reference to the reference method. This gives an operationally defined quantity.
- b) A second possibility is the use of two or more independent procedures to assess the procedure bias. If the results from independent procedures agree within their respective uncertainties, the producer may conclude that the values obtained are not significantly influenced by the individual procedures and hence the measurand is not operationally defined.
- c) In other cases, only a comparison among different laboratories using the same procedure is possible. In this case, it is impossible to demonstrate absence of method bias; therefore, the result is an operationally defined measurand.

**EXAMPLE** The mass fraction of Cd in soil was determined in several laboratories that all used aqua regia extraction and subsequent quantification by ICPMS. The measurand is operationally defined as “as obtained by aqua regia extraction and subsequent quantification by ICPMS”.

The definition of the property on the documentation provided to the users shall reflect the characterization approach chosen.

#### 9.2.5 Verification of traceability

In many cases, it is difficult to demonstrate the proper calibration of each and every piece of equipment. Such situations can arise because of unknown influence factors, but arise more frequently in characterization studies involving multiple laboratories, where obtaining calibration certificates for each and every instrument used is impractical. In these cases, the adequacy of the measures taken to ensure proper calibration of equipment and the traceability of results should be verified by, for example, specially designed and

vzorky (jako jsou vzorky jinak používané pro kalibraci) a certifikovanými referenčními materiály. Souhlas s výsledky vzorků řízení kvality se může používat jako důkaz dostačující kalibrace všech příslušných vstupních faktorů.

POZNÁMKA Důkaz shody s ISO/IEC 17025, včetně důkazů z posuzování třetí stranou, se může brát jako doplňkový důkaz návaznosti výsledků uváděných různými měřicími laboratořemi.

### 9.3 Charakterizace měřené veličiny nezávislé na postupu jednou metodou v jedné laboratoři

#### 9.3.1 Charakterizace referenční metodou bez přímého porovnání s CRM stejného druhu

##### 9.3.1.1 Koncepce

U toho přístupu se hodnota přiděluje v jediné laboratoři za použití jediné metody bez přímého porovnání s CRM podobného charakteru. Toto omezení počtu metod a laboratoří značně omezuje možnost detekovat neočekávané vlivy. Tento postup tudíž vyžaduje, aby byla k dispozici metoda, která je dostatečně dobře známa, aby se vyloučily nepoznané vlivy.

POZNÁMKA 1 „CRM stejného druhu“ znamená CRM, který odpovídá charakterizovanému CRM ve všech parametrech, které mohou ovlivnit výsledek měření (matrice, měřená vlastnost, hodnota veličiny měřené vlastnosti, atd.).

Výsledky a jejich nejistota se pak berou jako přidělená hodnota.

POZNÁMKA 2 Tento přístup je do značné míry ekvivalentní „charakterizaci jednou (primární) metodou v jediné laboratoři“ uváděné v Pokynu ISO 34:2009.<sup>[2]</sup>

##### 9.3.1.2 Požadavky na metodu

Každý měřicí postup, použitý v tomto přístupu, musí splňovat následující požadavky:

- je zcela pochopen, ve smyslu, že všechny kroky mají spolehlivý teoretický základ, takže systematická chyba je zanedbatelná ve vztahu zamýšlenému použití,
- je beze zbytku popsána rovnicí měření obsahující všechny relevantní ovlivňující faktory spojující měřenou veličinu a skutečně měřené vlastnosti, ze kterých jsou všechny vyjádřené v SI jednotkách,
- rovnice měření neobsahuje empiricky stanovené faktory, které mají hlavní vliv na výsledek měření (např. „míry výtěžnosti“).
- neexistuje žádný závažný vliv měřené veličiny na kterýkoliv ovlivňující faktor obsažený v rovnici,

prepared control samples (such as a sample otherwise used for calibration) and CRMs. Agreement of results on quality control samples can be used as demonstration of sufficient calibration of all relevant input factors.

NOTE Evidence of compliance with ISO/IEC 17025, including evidence from third party assessment, can be taken as additional evidence of traceability of the results reported by different measurement laboratories.

### 9.3 Characterization of a method-independent measurand by a single method in a single laboratory

#### 9.3.1 Characterization by a reference method without direct comparison with a CRM of the same kind

##### 9.3.1.1 Concept

In this approach, a value is assigned by one laboratory using only one method without direct comparison of a closely matched CRM. This limitation on the number of methods and laboratories greatly limits the possibility to detect unexpected effects. Therefore, this approach requires the availability of a method that is sufficiently well understood that unknown effects can be ruled out.

NOTE 1 “CRM of the same kind” refers to a CRM which matches the CRM to be characterized in all parameters that may have an influence on the measurement result (matrix, measured property, quantity value of the measured property, etc.).

The results and their uncertainty are then used as assigned value.

NOTE 2 This approach is largely equivalent to the “characterization by a single (primary) method in a single laboratory” in ISO Guide 34:2009.<sup>[2]</sup>

##### 9.3.1.2 Method requirements

Any measurement procedure used for this approach shall fulfil all of the following requirements:

- it is completely understood, meaning that all steps have a sound theoretical foundation so that systematic error is negligible relative to the intended use;
- it is completely described by a measurement equation containing all relevant influence factors linking the measurand to the properties actually measured, all of which can be expressed in SI Units;
- the measurement equation does not contain empirically determined factors that have a major influence on the measurement result (e.g. “recovery rates”);
- there is no relevant influence of the measured quantity on any of the influence factors contained in the equation;

- konstanty obsažené v rovnici jsou známy s nízkou nejistotou, která může být vyjádřena v SI jednotkách,
  - realistická bilance nejistoty může být vyjádřena v SI jednotkách a vycházet z individuální kvantifikace ovlivňujících faktorů zastoupených v rovnici,
  - nejistota měření výsledků získávaných metodou je malá ve srovnání s nejistotami obvykle dosahovanými při použití jiných metod pro předmětnou aplikaci.
- the constants contained in the equation are known with a low uncertainty, which can be expressed in SI units;
  - a realistic uncertainty budget can be written down in terms of SI units based on the individual quantification of the influence factors contained in the equation;
  - the measurement uncertainty of the results obtained by the method is small compared to the uncertainties usually obtained by application of other methods for the application in question.

I když není vyloženě nutně vyžadováno potvrzení nezávislou metodou, je nicméně velmi žádoucí poskytnout dodatečnou jistotu ve výsledky i když výsledky potvrzujících měření budou mít vyšší nejistotu než ty z referenční metody. Potvrzující metoda se může též použít k prokázání, že materiál je použitelný i pro jiné metody měření než je referenční metoda použitá pro charakterizaci.

While confirmation by an independent method is not strictly necessary, it is nevertheless highly advisable to provide additional confidence in the results, even if the confirmatory measurement results have a higher uncertainty than those from the reference method. The confirmatory method can also be used to demonstrate the applicability of the material to measurement methods other than the reference method used for characterization.

### 9.3.1.3 Měření pro charakterizaci

Pokud materiál sestává z mnoha jednotek (charakterizace šarže) měly se vybírat jednotky tak, aby pokryly celou šarži CRM. Jestliže materiál vykazuje významnou heterogenitu mezi jednotkami, mělo by se zkoušet více jednotek, aby se minimalizoval vliv heterogenity na výsledky.

Výrobce RM by měl navrhnout schéma měření (počet replikovaných výsledků, atd.), aby se dosáhlo cílové nejistoty. Kromě měření referenční metodou se velmi doporučuje provést potvrzující měření pomocí nezávislé metody, aby se vyloučily hrubé chyby.

### 9.3.1.3 Characterization measurements

If the material consists of multiple units (batch characterization), units should be selected covering the complete CRM batch. If the material shows significant between-unit heterogeneity, more units should be tested to minimize the effect of heterogeneity on the results.

The RM producer should propose a measurement scheme (number of replicate results, etc.) to achieve the target uncertainty. In addition to measurements with the reference method, it is highly recommended to perform confirmation measurements with an independent method to confirm absence of gross errors.

Přidělenou hodnotou je výsledek získaný referenční metodou. Standardní nejistota výsledku je vyjádřena jako  $u_{\text{char}}$ .

The assigned value is the result obtained by the reference method. The standard uncertainty of the result is expressed as  $u_{\text{char}}$ .

Výsledky potenciálních potvrzujících měření se nemusí slučovat s výsledky referenční metody, jelikož jejich nejistota je obvykle mnohem vyšší. Namísto toho se výsledky z různých metod testují, zda výsledky nezávislé metody souhlasí s těmi z referenční metody. Pokud se tomu tak, je to důkaz neexistence vychýlení (bias). Pokud tomu tak není, měla by se určit příčina (buď vychýlení potvrzující metody, nebo neočekávaný efekt u referenční metody) a výsledek korigovat, pokud je to nutné.

Results of potential confirmation measurements do not need to be combined with the results of the reference method, as their uncertainty is generally much higher. Instead, the results from the different methods are tested to determine whether the results of the independent method agree with those from the reference method. If this is the case, there is no evidence of method bias. If this is not the case, the cause (either a bias in the confirmation method or an unexpected effect in the reference method) should be identified and the result corrected, if necessary.

## 9.3.2 Charakterizace přímým porovnáním s CRM stejného druhu

## 9.3.2 Characterization by direct comparison with a CRM of the same kind

### 9.3.2.1 Princip

U tohoto přístupu se přiřazují hodnoty k „sekundárnímu CRM“ přímo porovnáním výsledků kandidátského CRM s výsledky již charakterizovaného a blízce podobného CRM („primární CRM“). Mezi příklady takových materiálů patří roztoky stopových prvků

### 9.3.2.1 Principle

In this approach, values are assigned to a “secondary CRM” by directly comparing results on the candidate CRM with those on an already characterized and closely matched CRM (the “primary CRM”). Examples for such materials include trace element solutions

měřené oproti certifikovaným roztokům, materiály měřené oproti standardům lékopisu nebo standardy absorbance měřené v porovnání s certifikovanými standardy absorbance.

**POZNÁMKA 1** Každé měření kandidátského referenčního materiálu, které vyžaduje kalibraci je ve skutečnosti porovnáním s jiným CRM (kalibrátorem). Tato kapitola se zabývá výhradně případem, kdy dva certifikované referenční materiály se tak podobají, že přímé porovnání v jedné laboratoři jedinou metodou může postačovat pro přidělení certifikované hodnoty.

U toho přístupu se nesmí sekundární CRM odlišovat od primárního CRM v žádném z parametrů, které mají významný vliv na výsledek měření. V tomto smyslu se musí zvažovat následující hlediska.

a) primární a sekundární certifikovaný materiál je stejné matrice. Malé rozdíly umožňující konstrukci kalibrační křivky jsou přijatelné, pokud hlavní charakteristiky materiálu zůstanou neměnné. Primární a sekundární certifikovaný materiál představují stejnou zátěž pro používanou metodu.

**PŘÍKLAD 1** Je možno charakterizovat roztok Cd v HNO<sub>3</sub> pomocí certifikovaného roztoku Cd v HNO<sub>3</sub>, protože matrice (zředěná HNO<sub>3</sub>) je u obou certifikovaných referenčních materiálů stejná. Není však možné charakterizovat Cd v granitu, jako sekundární CRM pomocí roztoku Cd v HNO<sub>3</sub>, protože se matrice liší. U mnoha metod přidává též krok rozkladu další analytický problém.

**PŘÍKLAD 2** Pokud chromatogramy sekundárních certifikovaných referenčních materiálů vykazují koeluci sledovaného analytu, není možné je charakterizovat jako sekundární certifikované referenční materiály. Pokud se primární a sekundární certifikované referenční materiály liší komplexitou (například mnohasložkové sekundární certifikované materiály charakterizované pomocí série nesmíchaných jednosložkových certifikovaných referenčních materiálů) výrobce by měl prokázat, že tato komplexita navíc neovlivňuje výsledek.

b) Pokud měřená veličina není definována postupem, zda pro předemná měření je daná matrice takového druhu, že měřicí postup může být považován za plně pochopený.

**PŘÍKLAD 3** Chromatografické stanovení roztoku benzo[a]pyrenu je dostatečně poznáno, protože se u něj nevyskytují žádné koeluce ani matricové vlivy. Stanovení benzo[a]pyrenu v zemině (nebo v extraktu ze zeminy) není plně zvládnuto, jelikož mnoho faktorů může ovlivňovat účinnost extrakce a může docházet k mnoha koelucím.

c) Kvantitativní úroveň měřené vlastnosti se dostatečně shoduje, aby se vyloučilo významné vychýlení způsobené různými úrovněmi.

**POZNÁMKA 2** U chemických měření tyto podmínky prakticky omezují výrobu sekundárních certifikovaných referenčních materiálů na čisté látky, roztoky/zředování čistých látek nebo vlastnosti definované postupem.

measured against certified solutions, materials measured against Pharmacopoeia standards or absorbance standards measured against certified absorbance standards.

**NOTE 1** Each measurement on a candidate CRM that requires calibration in fact compares it with another CRM (the calibrator). This clause deals entirely with the case where the two CRMs are so closely matched that a direct comparison in one laboratory using one method can be sufficient to assign a certified value.

For this approach, the secondary CRM shall be indistinguishable from the primary CRM in all parameters which have a significant influence on the measurement result. In this respect, the following aspects shall be considered.

a) The primary and secondary CRMs consist of the same matrix. Small differences to allow for the establishment of calibration curves are acceptable, as long as the main characteristics of the material remain unchanged. The primary and secondary CRMs present the same analytical challenges for the method used.

**EXAMPLE 1** It is possible to characterize a solution of Cd in HNO<sub>3</sub> against a certified solution of Cd in HNO<sub>3</sub>, as the matrix (diluted HNO<sub>3</sub>) of both CRMs is the same. It is not possible to characterize Cd in granite as a secondary CRM against a solution of Cd in HNO<sub>3</sub> as the matrix differs. For many methods, the digestion step also adds an additional analytical challenge.

**EXAMPLE 2** If chromatograms of secondary CRMs show co-elution at the analyte of interest, it is impossible to characterize them as secondary CRMs. If the primary and secondary CRMs differ in complexity (e.g. multi-component secondary CRMs characterized against a series of unmixed single-component primary CRMs) the producer should demonstrate that this added complexity does not influence the result.

b) If the measurand is not operationally defined, the matrix is of a kind that, for the measurement in question, the measurement procedure can be regarded as completely understood.

**EXAMPLE 3** The chromatographic determination of a solution of benzo[a]pyrene is sufficiently understood, as no co-elutions or matrix effects occur. Determination of the benzo[a]pyrene in soil (or, in a soil extract) is not fully understood, as a multitude of factors can influence extraction efficiency and many co-elutions can occur.

c) The quantity level of the measured property is sufficiently similar to rule out a significant bias due to the different levels.

**NOTE 2** For chemical measurements, these conditions practically restrict the production of secondary CRMs to pure substances, solutions/dilutions of pure substances or operationally defined properties.



Postup měření použitý k charakterizaci musí splňovat všechny kritéria návaznosti uvedená v Pokynu ISO 34:2009, 5.12,<sup>[2]</sup> a zaměřovat se na měřenou veličinu, pro kterou je charakterizován primární CRM.

Daný referenční měřicí postup musí prokázat platnost přenosu hodnoty z primárního na sekundární CRM.

### 9.3.2.2 Přidělená hodnota a $u_{\text{char}}$

Přidělená hodnota se musí vypočítat přímým srovnáním výsledků získaných pro primární a sekundární certifikované referenční materiály. Platnými metodami jsou metody s ohraničením (bracketing), několikabodové kalibrační křivky s primárním CRM, jednobodové kalibrace s primárním CRM o blízké hodnotě úrovně analytu.

$u_{\text{char}}$  vzniká sloučením nejistoty certifikované hodnoty primárního CRM, nejistoty kalibrace podle zvoleného kalibračního modelu a vlivu opakovatelnosti výsledků sekundárního CRM.

### 9.3.2.3 Návaznost

Certifikované hodnoty sekundárního CRM jsou návazné prostřednictvím primárního CRM ke stejné referenci jako hodnoty primárního CRM

**PŘÍKLAD** Roztok Cd v HNO<sub>3</sub> (sekundární CRM) se charakterizuje měřením oproti certifikovanému roztoku Cd v HNO<sub>3</sub> (primární CRM). Certifikované hodnoty primárního CRM jsou návazné na Mezinárodní soustavu jednotek. Tudiž certifikovaná hodnota sekundárního CRM je rovněž návazná na Mezinárodní soustavu jednotek.

### 9.3.3 Metody přípravy preparátů mísením

Tento přístup se obvykle používá pro přípravu kalibračních roztoků z čistých látek a též pro směsi plynů, o jejichž výrobě pojednává zvláštní norma.<sup>[30]</sup> Uvedený přístup se se někdy též používá pro výrobu maticových materiálů.

U všech mísených materiálů se musí znát hodnota měřené veličiny a její nejistoty. V mnoha případech se to rovná stanovení čistoty předmětného materiálu (viz 9.6) a potvrzení, že analyt není přítomen v materiálu, ke kterému má být přidán (například rozpouštědlo nebo 'blankového' maticového materiálu).

Je důležité dávat pozor na změny obsahu v průběhu přípravy a mísení, například by se měla v případě potřeby vyloučit ztráta vody nebo její absorpce.

Když se připravují směsi několika materiálů gravimetricky, každý z těchto materiálů se musí charakterizovat pomocí jednoho z přístupů popsaných v této kapitole.

Příprava pomocí volumetrie vychází ze stejných principů pro výpočet přidělené hodnoty a nejistoty, ale přináší dodatečnou potřebu věnovat zvýšenou pozornost neaditivitě objemů při míchání kapalin (například objemy směsí etanol/voda nejsou

The measurement procedure used for characterization shall fulfil all criteria for traceability listed in ISO Guide 34:2009, 5.12,<sup>[2]</sup> and address the measurand for which the primary CRM is characterized.

The **RMP** shall demonstrate the validity of the value transfer from the primary to the secondary CRM.

### 9.3.2.2 Assigned value and $u_{\text{char}}$

The assigned value shall be calculated by direct comparison between the results obtained on the primary and secondary CRMs. Valid methods include bracketing, multi-point calibration curves with the primary CRM, one point calibration with a primary CRM of closely matched analyte level.

$u_{\text{char}}$  consists of a combination of the uncertainty of the certified value of the primary CRM, the uncertainty of calibration according to the chosen calibration model, and the effect of the repeatability on the results of the secondary CRM.

### 9.3.2.3 Traceability

The certified values of the secondary CRM are traceable, via the primary CRM, to the same reference as the values of the primary CRM.

**EXAMPLE** A solution of Cd in HNO<sub>3</sub> (secondary CRM) is characterized by measurement against a certified solution of Cd in HNO<sub>3</sub> (primary CRM). The certified values of the primary CRM are traceable to the International System of Units. Therefore, the certified value of the secondary CRM is traceable to the International System of Units as well.

### 9.3.3 Formulation methods

This approach is usually applied for the preparation of calibration solutions from pure substances and also for gas mixtures, the production of which is described in a separate standard.<sup>[30]</sup> The approach is sometimes also used in the production of matrix materials.

The value of the measurand and its uncertainties, in all materials to be mixed, has to be known. In many cases, this is equivalent to determining the purity of the material of interest (see 9.6) and confirmation of the absence of the analyte in the material to which it is added (for example, a solvent or 'blank' matrix material).

It is important to guard against change in content between preparation and mixing; for example, water loss or uptake should be excluded, where appropriate.

If gravimetric mixtures of several materials, all of which contain the measurand in question, are to be prepared, each of the materials shall be characterized using one of the approaches described in this clause.

Volumetric preparation follows similar principles in calculation of assigned value and uncertainty but entails an additional need to pay close attention to non-additive volumes in mixing liquids (for example, ethanol/water mixture volumes are not the simple sum

jednoduchým součtem mísených objemů vody a etanolu) a dalším faktorům majícím vliv na měřený objem, zvláště teplotě.

V případě čistě gravimetrické přípravy pro certifikaci hmotnostního zlomku, se přidělená hodnota  $y_{\text{char}}$  vypočítá z hmotností jednotlivých komponent  $m_i$  a jejich hmotnostních zlomků analytu  $w_i$  jako

$$y_{\text{char}} = \frac{\sum w_i \cdot m_i}{\sum m_i} \quad (13)$$

U téhož postupu nejistota  $u_{\text{char}}$  zahrnuje všechny nejistoty kroků navažování a rovněž i nejistoty jednotlivých hmotnostních zlomků. Pro nejběžnější případ mísení dvou látek, je možno jí vypočítat pomocí:

$$u_{\text{char}} = \sqrt{\left(\frac{m_1}{m_1 + m_2}\right)^2 \cdot u_{w_1}^2 + \left(\frac{2w_1m_1 + 2w_1m_2}{(m_1 + m_2)^2}\right)^2 \cdot u_{m_1}^2 + \left(\frac{m_2}{m_1 + m_2}\right)^2 \cdot u_{w_2}^2 + \left(\frac{2w_2m_2 + 2w_2m_1}{(m_1 + m_2)^2}\right)^2 \cdot u_{m_2}^2} \quad (14)$$

Stejně jako u referenčních metod je záhodno zkontrolovat výsledek gravimetrické přípravy měřením, aby se zjistily všechny chyby v postupu výroby, i když v zásadě výsledek přípravy postačuje pro přidělení hodnoty.

POZNÁMKA Zevrubné pojednání o použití gravimetrie v analýze plynů lze nalézt v literatuře (viz například odkazy <sup>[31]</sup> a <sup>[32]</sup>).

#### 9.4 Charakterizace měřené veličiny nezávislé na postupu za použití dvou či více metod prokazatelné přesnosti v jedné či více laboratořích

##### 9.4.1 Koncepce

**9.4.1.1** Pro mnohé měřené veličiny neexistují referenční metody, které by poskytovaly přesné výsledky na vhodné úrovni nejistoty. V těchto případech je nutné najít další prostředky jak vylepšit spolehlivost přidělené hodnoty. Přístup popsáný v této kapitole používá řadu datových souborů, které byly získány použitím různých metod a/nebo v různých laboratořích, k tomu aby

- prokázal nepřítomnost významného vychýlení (bias) měřících postupů tím, že nezávislé postupy prokazatelně poskytují stejné výsledky,
- prokázal nepřítomnost významného vychýlení (bias) každé z laboratoří souhlasem jejich výsledků,
- zlepšil spolehlivost přidělené hodnoty zpřůměrováním výsledků a tak snížil vliv opakovatelnosti a znáhodňování a snížil vliv mezilaboratorní variability nebo variability mezi metodami.

**9.4.1.2** Koncepce stanovení hodnot vlastností RM nezávislých na metodě na základě shody mezi různými

of the water and ethanol mixed) and other factors affecting measured volume, particularly temperature.

For the case of purely gravimetric preparation to certify a mass fraction, the assigned value  $y_{\text{char}}$  is calculated from the masses of the individual components and their mass fractions of the analyte  $w_i$  as

$$y_{\text{char}} = \frac{\sum w_i \cdot m_i}{\sum m_i} \quad (13)$$

For the same procedure, the uncertainty  $u_{\text{char}}$  comprises all the uncertainties from the weighing steps as well as the uncertainties of the individual mass fractions. For the most common case of mixing two substances, this may be calculated from:

As for reference methods, although the result of the preparation is in principle sufficient for value assignment, it is highly advisable to check the result of the gravimetric preparation by measurement to detect any mistakes in the processing steps.

NOTE A comprehensive discussion of the use of gravimetry in gas analysis is found in the literature (see, for example, References<sup>[31]</sup> and<sup>[32]</sup>).

#### 9.4 Characterization of a method-independent measurand using two or more methods of demonstrable accuracy in one or more laboratories

##### 9.4.1 Concept

**9.4.1.1** For many measurands, no reference methods exist that provide accurate results at the appropriate level of uncertainty. In these cases, it is necessary to find other means of improving the reliability of the assigned value. The approach described in this clause uses a number of data sets, obtained using different methods and/or in different laboratories to

- demonstrate absence of significant bias in measurement procedures by showing that independent procedures yield the same results;
- demonstrate the absence of significant laboratory bias for each laboratory by agreement among results;
- improve the reliability of the assigned value by averaging results, thus reducing the effect of repeatability and randomizing and reducing the effect of between-laboratory or between-method variation.

**9.4.1.2** The concept of the determination of the method-independent property values of an RM based

metodami, potenciálně prováděnými v různých laboratořích, vychází přinejmenším ze dvou předpokladů:

- a) Existuje soubor postupů a/nebo laboratoří, jejichž prostřednictvím je možno stanovit charakteristiky RM a získat výsledky s přijatelnou přesností.
- b) U většiny přístupů vyhodnocování dat se předpokládá, že rozdíly mezi jednotlivými výsledky jak v rámci metod/laboratoří, tak i mezi nimi, jsou v podstatě náhodné bez ohledu na příčiny (například variabilita patřící měřicím postupům, personálu, vybavení atd.).

**9.4.1.3** Aby byl tento přístup platný, všechny výsledky z použitých měřicích postupů a/nebo účastnicích se laboratoří musí stanovovat stejnou měřenou veličinu a výsledky musí být návazné (viz 9.2.2) ke stejným příslušným stanoveným referencím. Tyto reference se obvykle pečlivě volí při návrhu mezilaboratorní studie pro získání hodnoty vlastnosti.

**POZNÁMKA 1** I na „úrovni doby“ mohou mezi laboratořemi existovat rozdíly ve výkonnostních charakteristikách měřicích postupů a rovněž i rozdíly ve velikosti nejistoty.

**POZNÁMKA 2** ISO 5725-3<sup>[14]</sup> vyjmenovává různé cíle mezilaboratorních porovnávání, mezi nimi validaci metody, zkoušení způsobilosti a charakterizaci referenčních materiálů. Záměr studie má významné důsledky na uspořádání a vyhodnocení těchto jednotlivých studií. Je tedy důležité mít cíl charakterizace na zřeteli a nesměšovat jej s jinými účely, i když je logisticky spojen s například úlohou zkoušení způsobilosti (viz dodatek A.3).

## 9.4.2 Návrh studie

V ideálním případě by měly být součástí studií s několika metodami metody s přinejmenším dvěma odlišnými měřicími principy. U mezilaboratorních studií s mnoha účastníky a volným výběrem metod by se mělo dbát o dobré zastoupení metod vhodných pro stanovení konkrétní charakteristiky.

Měla by se věnovat pozornost výběru kalibračního standardu, tj. zda by měl používat každý z účastníků standard podle svého výběru, nebo zda bude všem účastníkům poskytnut společný standard. Patříčná pozornost by se měla věnovat i čistotě používaných kalibrátorů.

Laboratoře se musí vybírat na základě své prokázané kompetence. Tedy účastníci se laboratoře musí poskytnout důkaz kompetence pro předmětnou měřenou veličinu nezávisle na měřeních kandidátského CRM, ideálně před zahájením studie. Je tedy nepřijatelné používat data o kandidátském CRM ze stejné studie jako prokázání způsobilosti a pro přidělení hodnoty CRM (např. používat konsenzuální

on agreement among different methods, potentially performed in different laboratories, is based on at least two assumptions:

- a) There exists a population of procedures and/or laboratories that is capable of determining the characteristics of the RM and providing results with acceptable accuracy.
- b) For most data evaluation approaches, it is assumed that the differences between individual results, both within and between methods/laboratories, are random in nature regardless of the causes (for example, variation in measurement procedures, personnel, equipment, etc.).

**9.4.1.3** For this approach to be valid, all results of all measurement procedures and/or laboratories involved shall determine the same measurand and the results shall be traceable (see 9.2.2) to the same, appropriate stated references. These references are usually carefully selected when designing a collaborative study to obtain the property value.

**NOTE 1** Even at the “state-of-the-art” level, differences in performance characteristics of measurement procedures as well as differences in the size of uncertainty can exist between laboratories.

**NOTE 2** ISO 5725-3<sup>[14]</sup> lists various objectives of interlaboratory comparisons, among them method validation, proficiency testing and characterization of reference materials. The goal of the study has important implications for the setup and evaluation of the various studies. It is therefore important to keep the goal of characterization in mind and not to mix it with other purposes, even if it is logistically combined with, for example, a proficiency testing exercise (see Clause A.3).

## 9.4.2 Study design

Ideally, at least two substantially different measurement principles should be included in a multiple-method study. For interlaboratory studies using many participants with free choice of methods, a good representation of methods suitable for the determination of the particular characteristic should be sought.

Consideration should be given to the choice of the calibration standard, i.e. whether each participant should use a standard of its choice or whether a common calibrant is provided to all participants. The purity of the calibrants used shall be given due consideration.

Laboratories shall be selected based on demonstrated competence. Therefore, participating laboratories shall provide evidence of competence for the measurand in question independent of the measurements on the candidate CRM, ideally before commencement of the study. It is thus impossible to use data on the candidate CRM from the same study as demonstration of competence and for value assignment of a CRM (e.g. using the consensus value of results of a proficiency

hodnotu výsledků studie zkoušení způsobilosti pro přidělení hodnoty CRM) (viz též dodatek A.3).

Výrobce by měl sestavit dokumentovanou minimální řadu technicky platných výsledků, které se budou brát v úvahu pro přidělení hodnoty. Počet souborů dat by měl být dostatečně vysoký, aby poskytoval dostatečně malou nejistotu odhadované hodnoty s přihlédnutím k možným nedostatkům při uvádění, vyloučení výsledků z technických důvodů a možnostem zamýšleného statistického vyhodnocení.

**POZNÁMKA 1** Počet účastnících se laboratoří je méně významný než počet nezávislých datových souborů. Jednotlivá laboratoř může být schopna dodat několik datových souborů, všechny získané nezávislými postupy a/nebo instrumentací.

Organizátor musí začlenit odpovídající opatření řízení kvality, aby zajistil kvalitu dodávaných výsledků.

Výrobce by měl specifikovat formu zprávy, která může představovat předem naformátované formuláře pro nahlášení údajů účastníky. Zprávy by měly obsahovat dostatek podrobností pro možnou kontrolu technické platnosti výsledků.

**POZNÁMKA 2** Když se zprávy doručují ve formátu tabulkových procesorů, neúmyslné změně lze často zabránit použitím „zamknutím“ nebo „zajištěním“ prostředků, použitých v tabulkovém procesoru.

Organizátor by měl poskytnout dostatek pokynů účastnícím se laboratořím a/nebo operátorům, aby se zajistilo hladké uskutečnění prací. Aby byla úspěšná, musí mít mezilaboratorní studie přesně stanovený cíl, musí být navržena efektivně, účinně organizována s jasnými stručnými pokyny, kterým jsou všichni zúčastnění schopni snadno dostát. Účast v takovém programu, ať již jako operátor či laboratoř znamená souhlas s tím, že se budou řídit těmito pokyny.

**POZNÁMKA 3** Další směrnice pro organizaci studií s mnoha metodami v jedné či více laboratořích obsahuje příloha A.

## 9.4.2 Vyhodnocení

### 9.4.3.1 Technické a statistické vyhodnocení

Soubory dat by se měly prozkoumat vizuálně a graficky. Data poskytnutá každou laboratoří by se měla zkontrolovat, zda jsou kompletní a jakákoliv zjištěná anomálie se musí pečlivě zkontrolovat na možnost triviálních (chyby přenosu, chyby tisku atd.) a netriviálních (výpadek, selhání vybavení, atd.) příčin. Pokud se chyby či poruchy potvrdí, musí se odpovídající výsledky opravit nebo vyloučit.

Všechny výsledky se musí zkontrolovat, zda u nich nedošlo k technickým chybám podle informací o měřicích postupech, které poskytli účastníci studie. Technické zhodnocení musí poskytnout soubor technicky platných dat, tj. dat, které brány jednotlivě, lze považovat za nevyhýlený odhad pravé hodnoty.

test study for value assignment of a CRM) (see also Clause A.3).

The producer should set a documented minimum number of technically valid results for which value assignment will be considered. The number of data sets should be large enough to provide an adequately small uncertainty in the estimated value after allowing for the possibility of failure to report, exclusion of results for technical reasons and the intended statistical evaluation.

**NOTE 1** The number of participating laboratories is less important than the number of independent data sets. A single laboratory may be able to provide several data sets, all obtained by independent procedures and/or instruments.

The organizer shall implement adequate quality control measures to ascertain the quality of the results delivered.

The producer should specify the form of reporting, which may include preformatted reporting forms for participants. The reports should contain sufficient detail to check the technical validity of results.

**NOTE 2** When reports are submitted in spreadsheet form, unintentional alteration can often be prevented by the use of 'locking' or 'protection' facilities incorporated in the spreadsheet software.

The organizer should provide sufficient guidance for participating laboratories and/or operators to ensure the smooth implementation of the work. To be successful, the collaborative study shall have a well-defined objective, be effectively designed and efficiently organized with clear, concise guidelines with which all involved can readily comply. Participation, either as operator or as laboratory, in such a program implies agreement to adhere to these guidelines.

**NOTE 3** Additional guidance on the organization of multiple-method studies in one or more laboratories is given in Annex A.

## 9.4.2 Evaluation

### 9.4.3.1 Technical and statistical evaluation

Data sets should be inspected visually and graphically. Data submitted by each laboratory shall be checked for completeness and any observed anomaly shall be examined carefully for possible trivial (transmission error, misprint, etc.) and non-trivial reasons (drop-out, equipment failure, etc.). If errors or failures are confirmed, the corresponding results shall be corrected or rejected.

All results shall be checked for the absence of technical errors based on the information on the measurement procedures provided by the study participants. The technical evaluation shall lead to a set of technically valid data, i.e. data that each taken alone would be regarded as an unbiased estimate of the true value.

POZNÁMKA Zařazení materiálů pro řízení kvality, se známými hodnotami, do takových studií se ukázalo neocenitelné pro identifikaci technických problémů.

Spojená množina technicky přijatelných souborů dat musí být vyhodnocena statisticky, se zřetelem na výchozí rozdělení hodnot, přítomnost shluků výsledků a potenciálních odlehklých hodnot. Pro daný soubor dat se musí zvolit vhodné statistické metody a parametry, které se budou certifikovat.

Všechna vyhodnocení by měla brát v úvahu nejistoty měření udávaných výsledků.

#### 9.4.3.2 Přidělená hodnota

Přidělená hodnota se obvykle vypočítá jako vážený nebo nevážený průměr nebo nejpravděpodobnější hodnota (průměr, medián, modus, atd.). Pokud se použijí váhící faktory, měl by výrobce poskytnout doklad pro použitý faktor.

Tam, kde průměry datových souborů vykazují normální rozdělení a nepoužije se vážení, nevážený aritmetický průměr z průměrů  $y_i$   $p$  datových souborů se použije jako přidělená hodnota  $y_{\text{char}}$ .

$$y_{\text{char}} = \frac{\sum y_i}{p} \quad (15)$$

Průměrná hodnota jednotlivých výsledků se může též použít jako přidělená hodnota tam, kde rozdíly mezi průměry datových souborů jsou nevýznamné ve srovnání s projevy variability uvnitř každého souboru dat.

Nezbytným předpokladem pro použití vážené hodnoty je dostupnost spolehlivých odhadů nejistoty u všech hodnot, protože nejistoty přidružené k váženým odhadům jsou často ve větší míře ovlivňovány nesprávnými jednotlivými nejistotami než nevážené odhady. Různé přístupy k vážení popisují například odkazy [33] a [34]. Například vážení podle nejistot může zavést vychýlení nebo nepříznivě ovlivňovat rozptyl použitého odhadovaného faktoru, pokud všechny nejistoty nebyly spolehlivě odhadnuty. Vážení podle zjištěných rozptylů má za nedostatek to, že pozorované rozptyly mohou často zohledňovat pouze část reálné nejistoty měření.

Pokud výsledky nevykazují normální rozdělení a nemohou být transformovány na data s normálním rozdělením, použijte statisticky platný přístup, který odpovídá pozorovanému rozdělení.

#### 9.4.3.3 Odhad $u_{\text{char}}$

Nejistotu charakterizace lze odhadnout na základě údajů o nejistotě, které poskytla laboratoř nebo z dodaných dat bez přihlídnutí k údajům o nejistotě dodané laboratoří nebo kombinací obou přístupů.

#### Vyhodnocení z nejistot

NOTE Inclusion of quality control materials, with known values, in such studies has been found invaluable to identify technical problems.

The pool of technically accepted data sets shall be evaluated statistically, giving due consideration of the underlying distribution of values, presence of clusters of results and potential outliers. Appropriate statistical methods for the data set and parameter to be certified shall be selected.

All evaluations should take the measurement uncertainties of the reported results into consideration.

#### 9.4.3.2 Assigned value

The assigned value is usually calculated as a weighted or unweighted mean or most likely value (mean, median, mode, etc.). If weighting factors are used, the producer should provide evidence for the validity of the estimator used.

Where the data set means follow approximately a normal distribution and no weighting is applied, the unweighted arithmetic mean of the  $p$  data set means  $y_i$  is applied as assigned value  $y_{\text{char}}$ .

$$y_{\text{char}} = \frac{\sum y_i}{p} \quad (15)$$

The mean value of individual results may also be adopted as the assigned value where differences between data set means are insignificant compared to the effects of variation within each data set.

A prerequisite for the use of a weighted value is the availability of reliable uncertainty estimates for all values, as uncertainties associated with weighted estimates are often much more seriously affected by incorrect individual uncertainties than unweighted estimates. Different approaches to weighting are described in, for example, References [33] and [34]. For example, weighting according to measurement uncertainties can introduce a bias or adversely affect the variance of the estimator used if not all uncertainties were estimated reliably. Weighting according to observed variances has the drawback that observed variances often account for only a part of the real measurement uncertainty.

If the results do not follow a normal distribution and cannot be transformed into normally distributed data, apply a statistically sound approach that is consistent with the observed distribution.

#### 9.4.3.3 Estimation of $u_{\text{char}}$

The uncertainty of characterization can be estimated either by using the uncertainty statements submitted by the laboratory or from the submitted data ignoring the uncertainty statements made by the laboratory or from the combination of both.

#### Uncertainty-based evaluation

Je teoreticky možné slučovat výsledky včetně jejich nejistot do jedné hodnoty (hodnoty vlastnosti) a kombinované standardní nejistoty.

Mezi popisované patří vážení výsledků podle nejistot<sup>[35]</sup>, určení detailních výrazů pro nejistoty<sup>[34]</sup>  $\chi^2$ -fitování,<sup>[36]</sup> nebo rozdělení nejistot na společné a individuální součásti.<sup>[37]</sup> Protože realizace těchto přístupů je pracná a velmi citlivá na pochybení, zjistilo se, že se těžce uplatňují a dosud se používají zřídka.

**POZNÁMKA** V literatuře se často  $\chi^2$ -fitování uvádí jako „nejmenší čtverce“ nebo „fitování pomocí nejmenších čtverců“. V podstatě  $\chi^2$ -fitování se liší od „fitování pomocí nejmenších čtverců“ jen faktorem měřítka a umožňuje vyhodnocení statistiky  $\chi^2$ -následně po fitování.

#### Vyhodnocení nejistot z laboratoří

Tam, kde průměry souborů data vykazují přibližně normální rozdělení a nepoužije se žádné vážení, směrodatná odchylka průměru  $p$  průměrů datových souborů  $y_i$  se může použít jako  $u_{\text{char}}$ .

$$u_{\text{char}} = \frac{s(y)}{\sqrt{p}} = \frac{1}{\sqrt{p}} \cdot \sqrt{\frac{\sum (y_i - y_{\text{char}})^2}{p - 1}} \quad (16)$$

kde  $s(y)$  představuje směrodatnou odchylku  $p$  hodnot průměrů datových souborů.

Pokud výsledky nevykazují normální rozdělení a nemohou být transformovány na data s normálním rozdělením, musí se použít metrologicky a statisticky platný přístup, který odpovídá pozorovanému rozdělení.

#### **9.4.4 Studie v jedné laboratoři využívající několik metod**

V některých případech organizace vynaložily mimořádné úsilí do vývoje metod, na metrologické řízení těchto metod, realizaci přístupů jako u referenčních metod. V takových případech datové soubory pouze z několika takových metod, jestliže jsou jejich principy měření dostatečně odlišné, mohou postačovat pro charakterizaci. Za těchto okolností:

- výrobce referenčních materiálů má pravděpodobně přístup ke kompletním údajům prokazování kvality a validačním údajům, které je možno brát v úvahu při technickém hodnocení,
- počet datových souborů je malý a vyhodnocení by se měla opírat o posouzení a použití nejistot měření přidružených ke každé metodě. Větší důraz je tedy třeba klást na posouzení nejistot měření a velký význam má soulad mezi odhadovanými nejistotami.

**POZNÁMKA** Jeden z nejčastěji používaných přístupů popsali Mandel and Paule.<sup>[38]</sup>

It is theoretically possible to combine the results, including their uncertainties, into a single value (the property value) and a combined standard uncertainty.

Approaches described include weighting results by uncertainties,<sup>[35]</sup> determining detailed expressions for the uncertainties,<sup>[34]</sup>  $\chi^2$ -fitting,<sup>[36]</sup> or splitting uncertainties into common and individual parts.<sup>[37]</sup> As implementing these approaches is laborious and quite sensitive to mistakes, they have been found difficult to implement and have so far rarely been used.

**NOTE** In the literature,  $\chi^2$ -fitting is also known as “least squares” or “least squares fitting”. In principle,  $\chi^2$ -fitting differs only from “least squares fitting” by a scaling factor, allowing evaluation of the statistic  $\chi^2$  after the fitting procedure.

#### Evaluation without the laboratories' uncertainties

Where the data set means follow an approximately normal distribution and no weighting is applied, the standard deviation of the mean of the  $p$  data set means  $y_i$  can be applied as  $u_{\text{char}}$ .

$$u_{\text{char}} = \frac{s(y)}{\sqrt{p}} = \frac{1}{\sqrt{p}} \cdot \sqrt{\frac{\sum (y_i - y_{\text{char}})^2}{p - 1}} \quad (16)$$

where  $s(y)$  denote the standard deviation of the  $p$  data set mean values.

If the results do not follow a normal distribution and cannot be transformed into normally distributed data, a metrologically and statistically sound approach that is consistent with the observed distribution shall be applied.

#### **9.4.4 Single-laboratory multi-method studies**

In some cases, organisations have invested an exceptional amount of effort in method development, such that the metrological control of these methods, approaches that of reference methods. In such cases, data sets from only a few of these methods, given that their measurement principles are sufficiently different, may be sufficient for characterization. Under these circumstances:

- the RMP is likely to have access to the complete quality assurance and validation data, which should be taken into consideration for the technical evaluation;
- the number of data sets is small and evaluation should rely on the assessment and use of measurement uncertainties associated with each method. More emphasis therefore needs to be put on the assessment of measurement uncertainties and agreements within the estimated uncertainties is of high importance.

**NOTE** One of the most widely used approaches for characterization was described by Mandel and Paule.<sup>[38]</sup>

## 9.5 Charakterizace měřené veličiny definované postupem v síti kompetentních laboratoří

### 9.5.1 Koncepce

Tento přístup lze použít pro výrobu referenčních materiálů certifikovaných na parametry definované postupem. Protože v takovém případě metoda definuje měřenou veličinu, prokázání neexistence vychýlení laboratoře je často možné jen na základě sdružování údajů z několika laboratoří. Kromě toho je definující metoda obvykle relativně nepřesná a jediným praktickým způsobem jak dosáhnout malou nejistotu je zprůměrovat mnoho výsledků z různých laboratoří.

Tento přístup je do značné míry podobný přístupu popsanému v odstavci 9.4, s tou výjimkou, že všechny laboratoře používají stejný postup.

Předpokladem je opět existence několika laboratoří, které dovedou provádět stejně dobře předemné měření. Tento přístup se zaměřuje na znáhodnění ovlivňujících faktorů v rámci mezí stanovených v metodě.

### 9.5.2 Uspořádání studie

Musí se zvolit dobře popsaný postup měření. V mnoha případech se bude jednat o mezinárodně uznávaný postup (např. ISO, ASTM, AOAC, IFCC, atd.). Jakákoliv modifikace takového postupu odsouhlasená všemi účastníky (např. těsnější specifikace některých parametrů) má v principu za následek modifikaci postupu a identita měřené veličiny bude pak definována odkazem na modifikovanou metodu.

**POZNÁMKA 1** Předběžné studie mohou ukázat na neúmyslné odchylky od standardního postupu, které by měly být pak korigovány před přistoupením k vlastní charakterizaci.

Na výběr laboratoří a přijímaná opatření řízení kvality se vztahují pokyny popsané v odstavci 9.4.2.

**POZNÁMKA 2** V tomto případě se mohou též používat vzorky na řízení kvality, aby se prokázalo, že konkrétní přístroj splňuje všechny specifikace.

Měl by se vybrat takový počet laboratoří, aby se získala dostatečně malá nejistota pro zamýšlené použití daného RM.

### 9.5.3 Vyhodnocení

V případě měřených veličin definovaných postupem je definující metoda (podle definice) nezátížená vychýlením a při vyhodnocování nejistoty je nutné pouze zvažovat možné vychýlení laboratoře a vlivy uvnitř laboratoře.

Na vyhodnocení výsledků se vztahují přístupy popsané v dodatku A.2.

## 9.5 Characterization of an operationally defined measurand using a network of competent laboratories

### 9.5.1 Concept

This approach is applicable to the production of RMs certified for operationally defined parameters. As in this case the method defines the measurand, demonstration of absence of a laboratory bias is often only possible by combining data from several laboratories. In addition, the defining method is often relatively imprecise and the only practical means of obtaining a small uncertainty is to average many results from different laboratories.

This approach is largely similar with that described in 9.4, with the exception that all laboratories apply the same procedure.

The assumption is again that a number of laboratories exists that can perform the measurement in question equally well. The approach aims at randomization of all influence factors within the limits set by the method.

### 9.5.2 Study setup

A well-described measurement procedure shall be chosen. In many cases, this will be an internationally agreed procedure (e.g. ISO, ASTM, AOAC, IFCC, etc.). Any modification of such a procedure agreed by all participants (e.g. tighter specifications for some parameters) result in principle in a modified procedure and the measurand's identity is then defined by reference to the modified method.

**NOTE 1** Preliminary studies may show unintended departures from the standard procedure, which should be corrected before proceeding to characterization.

The considerations described in 9.4.2 apply for the selection of laboratories and quality control measures taken.

**NOTE 2** Quality control samples can also be used in this case to demonstrate that a particular instrument fulfils all specifications.

The number of laboratories should be chosen to provide a sufficiently small uncertainty for the intended use of the RM.

### 9.5.3 Evaluation

In the case of operationally defined measurands, the defining method is, (by definition), unbiased and it is then necessary only to consider possible laboratory bias and within-laboratory effects in an uncertainty evaluation.

The approaches described in Clause A.2 apply for the evaluation of results.

## 9.6 Čistota

### 9.6.1 Obecně

Čisté látky představují primární měřicí standardy (etalony) a základní zdroj metrologické návaznosti vyššího řádu pro většinu metrologických řetězců v chemii, termometrii a kalorimetrii obecně a pro certifikaci roztoků a maticových referenčních materiálů zvláště. Adjektivum „čistý“ se vztahuje k idealizované situaci: žádná látka není čistá ze 100 %, vždy bude obsahovat nečistoty na určité úrovni. Příslušná certifikace látek na čistotu je tedy neodmyslitelným základním kamenem návaznosti pro chemická měření.

Čisté látky jsou významnou třídou certifikovaných referenčních materiálů sami o sobě. Používají se v laboratořích buď k rozšiřování návaznosti vyšších řádů na kalibrační standardy používané u měřících metod, nebo při certifikaci a výrobě dalších certifikovaných referenčních materiálů, jako jsou například roztoky nebo směsi plynů.

Čistotu látek lze určit buď přímo (měřením předmětného analytu) nebo nepřímo odečtením hmotnosti či molárního zlomku všech nečistot od 100 %.

Když se charakterizuje čistota materiálu, měla by se dodatečně potvrdit jeho identita.

### 9.6.2 Přímé stanovení čistoty

V některých případech hmotnost či molární zlomek předmětné látky se může stanovit přímo. Mezi vhodné metody patří coulometrie, titrace a kalorimetrie (snížení bodu tuhnutí). Pro organické analyty se ve zvýšené míře zavádí pro přímou certifikaci čistoty referenčních materiálů technika kvantitativního NMR.<sup>[39] [40]</sup>

Metody vyžadující kalibraci předmětnou látkou (například HPLC, GC, ICP-MS, AAS, atd.) se mohou v podstatě použít pro stanovení čistoty, ale jedná se o sekundární měřicí metody. Jelikož vyžadují standard o známé čistotě, použití těchto metod pro přímé stanovení čistoty se často omezuje na přiřazení hodnot pracovním standardům.

Čistota stanovená takovou metodou je přijata jako přidělená hodnota a nejistota stanovení čistoty se použije jako  $u_{\text{char}}$ .

Potvrzení takových hodnot nezávislými měřeními se velmi doporučuje.

### 9.6.3 Nepřímé stanovení čistoty

Čistotu lze stanovit jako rozdíl, použitím souboru ortogonálních analytických technik schopných detekovat a kvantifikovat všechny hlavní třídy nečistot v materiálu následujícím způsobem:

## 9.6 Purity

### 9.6.1 General

Pure substances constitute the primary measurement standard and ultimate source of higher-order metrological traceability for most traceability chains in chemistry, thermometry and calorimetry in general and for the certification of solution and matrix reference materials in particular. The adjective “pure” refers to an idealised situation: no substance is 100 % pure, there will always be impurities present at some level. The appropriate certification of substances for purity is thus an essential cornerstone of traceability in chemical measurement.

Pure substances are an important class of CRMs in their own right. They are used by laboratories either to disseminate higher order traceability to calibration standards used in measurement methods, or in the certification and preparation of other CRMs, such as, for example, solutions or gas mixtures.

The purity of substances can either be determined directly (measuring the analyte in question) or indirectly by subtracting the mass or mole fractions of all impurities from 100 %.

When characterizing the purity of a material, the identity of the material should additionally be confirmed.

### 9.6.2 Direct determination of purity

In some cases, the mass or mole fraction of the substance in question can be determined directly. Suitable methods can include coulometry, titrimetry and calorimetry (freezing point depression). In the case of organic analytes the use of the technique of quantitative NMR for the direct certification of the purity of reference materials is increasingly being implemented.<sup>[39] [40]</sup>

Methods requiring calibration with the substance in question (e.g. HPLC, GC, ICP-MS, AAS, etc.) can in principle be used for purity assignment, but they are secondary measurement methods. Since they require a standard of known purity, the application of such methods for direct determination of purity is often limited to the assignment of values to working standards.

The purity determined by the method is adopted as assigned value and the uncertainty of the purity determination is adopted as  $u_{\text{char}}$ .

Confirmation of these values by independent measurements is highly recommended.

### 9.6.3 Indirect determination of purity

Purity can be determined by difference, using a set of orthogonal analytical techniques capable of detecting and quantifying all the major classes of impurities in the material, as follows:



- a) možné nečistoty se vyšetřují v odpovídajícím rozsahu, který vždy zahrnuje organická zbytková rozpouštědla, vodu, anorganické a organické nečistoty. Druhy zkoumaných nečistot často vyplývají z procesu výroby dané látky,
- b) v látce, která je předmětem certifikace, se stanoví množství každé z možných nečistot,
- c) čistota hlavní složky se vypočítá jako rozdíl.

Měření, která jsou nutná pro stanovení nečistot, mohou být náročná, jelikož většina nečistot bude přítomna blízko mezím detekce a stanovitelnosti, nebo budou těžce oddělitelné v případě nečistot blízko strukturně podobných ke hlavní složce. Dále různé měřicí metody mohou poskytovat výsledky v neslučitelných jednotkách (hmotnostní zlomky pro těkavé látky, molární zlomky pro celkové nečistoty z kalorimetrie), což činí slučování takových výsledků v přísně metrologickém smyslu nemožné, pokud nejsou známy struktury nečistot. Avšak pokud jsou úroveň nečistot nízké a je použita příslušná korekce na jejich molekulovou hmotnost aproximace hmotnostní zlomek = molární zlomek = podíl chromatografické plochy nezavede nepřiměřeně velké vychýlení (bias). Pokud jsou přítomny specifické nečistoty na vyšších úrovních, je nutná jejich identifikace pro oprávnění činit takové předpoklady. Kvantifikace každé nečistoty vůči specifickému kalibrantu je ideální, ale může být nepraktická nebo nemožná, pokud nejsou k dispozici dostatečné prostředky nebo příslušné referenční materiály nejsou dostupné. V takovém případě se musí začlenit příslušný příspěvek k dané nejistotě jako důsledek předpokladů o identitě a faktorech odezvy u individuálních nečistot.<sup>[41]</sup>

**POZNÁMKA** ICH harmonizovaná tripartitní směrnice Q3A "Nečistoty v nových léčivých látkách" <sup>[42]</sup> vyžaduje, aby se identifikovaly nečistoty mezi 0,05 and 0,1 % (v závislosti na denním přísunu dané látky).

Ačkoliv se mohou obdržet relativně velké nejistoty při kvantifikaci jednotlivých nečistot, pokud je absolutní úroveň nízká, příspěvek k nejistotě finální hodnoty hlavní složky je obvykle nízký.

Vyhodnocení nejistot se může komplikovat těsnou blízkostí matematických mezí (látkové množství a hmotnostní zlomek jsou definovány pouze mezi 0 a 1), což může vytvářet další problémy, včetně odhadů příspěvků některých tříd nečistot, které nabydou nominálně negativních hodnot.

Model pro certifikovanou hodnotu  $y_{\text{char}}$  jako látkové množství nebo hmotnostní zlomek hlavní komponenty  $y$  jako funkce  $k$  nečistot o látkovém množství nebo hmotnostních zlomcích  $w_i$  se rovná

$$y_{\text{char}} = 1 - \sum w_i \quad (17)$$

Za předpokladu nezávislosti mezi měřeními hmotnostních zlomků nečistot (což je častý případ),

- a) a suitable range of possible impurities are investigated, always including residual organic solvents, water, inorganic and organic impurities. The types of impurities to be investigated is often informed by the manufacturing process for the substance;
- b) the amount of each of the possible impurities is determined in the substance to be certified;
- c) the purity of the main component is computed by difference.

The measurements necessary to determine the impurities may be challenging, since most impurities will be close to the detection and/or determination limits or difficult to resolve in the case of impurities closely related in structure to the main component. Furthermore, different measurement methods may give results in non-compatible units (mass fractions for volatiles; mole fractions for total impurities by calorimetry), which makes the combination of such results in a strict metrological sense impossible if the structures of the impurities are not known. However, as long as the level of impurities is small and appropriate correction is made for their molecular weight, the approximation, mass fraction = mole fraction = chromatographic area fraction will not introduce an excessively large bias. If higher levels of specific impurities are present, their identification is necessary to justify such assumptions. Quantification of each impurity against specific calibrants is ideal, but may be impractical or impossible if sufficient resources or appropriate reference materials are not available. In this case, appropriate allowance has to be made for the uncertainty introduced as a result of assumptions regarding the identity and response factors of individual impurities.<sup>[41]</sup>

**NOTE** The ICH harmonized tripartite guideline Q3A "Impurities in new drug substances" <sup>[42]</sup> requires that impurities between 0,05 and 0,1 % (depending on the daily uptake of the substance) be identified.

Although high relative uncertainties may be obtained for the quantification of individual impurities, provided the absolute level is small, the contribution to the uncertainty of the final value for the main component is usually low.

The evaluation of the uncertainties can also be complicated by the proximity of mathematical limits (amount of substance and mass fractions are only defined between 0 and 1), which may create additional problems, including estimates for some contributing impurity classes that include nominally negative values.

The model for the certified value  $y_{\text{char}}$  of the amount of substance or mass fraction of the main component  $y$  as a function of  $k$  impurities with amount of substance or mass fractions  $w_i$  is

$$y_{\text{char}} = 1 - \sum w_i \quad (17)$$

Assuming independence among measurements of the mass fractions of the impurities (which is often the

kombinovaná standardní nejistota přidružená k látkovému množství nebo hmotnostnímu zlomku hlavní složky se rovná

$$u_{\text{char}}^2 = \sum u^2(w_i) \quad (18)$$

Často se stává, že některé zlomky látkového množství nebo hmotnostní zlomky  $w_i$  se rovnají nule, vzhledem ke skutečnosti, že žádná z těchto nečistot se opravdu nevyskytuje, nebo jejich úroveň je pod mezí detekce měřicí metody. Tam, kde hodnota pro nečistotu je pod mezí detekce, udává se obecně jako nula a mez detekce metody se použije ke stanovení standardní nejistoty přidružené k přidělené nulové hodnotě.

## 9.7 Identita

### 9.7.1 Materiály s certifikací podle původu

Při charakterizaci referenčního materiálu se může vycházet ze znalosti původu materiálu, tedy jeho provenience.

Aby se mohla charakterizace opírat o provenienci materiálu, měl by výrobce získat dokumentaci nebo jiný důkaz původu materiálu, který prokazuje nepřerušovaný řetězec od zdroje k finálnímu balení. Taková dokumentace by se měla uchovávat po celou dobu životnosti daného materiálu.

Výrobci by měli mít na místě postupy, které by zajišťovali, že zacházení s materiálem (včetně vzorkování, homogenizace, balení, skladování atd.) zabraňuje kontaminaci jinými materiály a nemění odezvu typických zkušebních metod, pro které je materiál určený.

Tam, kde je to možné by měl výrobce podniknout experimentální ověření (včetně měření, odbornou inspekci nebo kvalitativní testování – podle okolností), aby se potvrdila identita přiřazená podle provenience.

**PŘÍKLAD** DNA extrahovaná z bakteriální kultury vzniklé množением z jediné bakterie, která byla získána z banky referenčního kmene <sup>[43]</sup> mohla být certifikována na základě provenience s podmínkou konfirmačních kontrol na kontaminaci.

### 9.7.2 Materiály certifikované na identitu prokázanou měřením

#### 9.7.2.1 Obecně

Když se charakterizuje identita látek na základě měření, je třeba mít na mysli několik hledisek.

a) Identita není obvykle výsledkem měření, ale závěrem učiněným na základě výsledků měření jedné či několika metod. Například chemické posuny a výšky píků v NMR spektru, nebo kombinace barvy, bodu tání, molární hmotnosti atd. mohou udávat charakter přiřazení identity. Zatímco nejistoty měření mohou být přiřazeny individuálním výsledkům měření, jejich slučování, aby poskytly nějakou numerickou indikaci nejistoty identity

case), the combined standard uncertainty associated with the amount-of-substance or mass fraction of the main component is

$$u_{\text{char}}^2 = \sum u^2(w_i) \quad (18)$$

It frequently happens that some of the impurity amount of substance fractions or mass fractions  $w_i$  are zero, due to the fact that either these impurities are truly absent, or that their levels are below the detection limit of the measurement method. Where a value for an impurity is below the detection limit, it is generally reported as zero and the method detection limit used to establish the standard uncertainty associated with the assigned zero value.

## 9.7 Identity

### 9.7.1 Materials certified based on provenance

A reference material may be characterized based on knowledge of the origin of the material, that is, the provenance of the material.

To support characterization based on provenance, the producer should obtain documentary or other evidence of the origin of the material that shows an unbroken chain of evidence from origin to final packaging. The documentation should be maintained for the lifetime of the material.

Producers should have procedures in place to ensure that handling of the material (including sampling, homogenization, packaging, storage, etc.) prevents contamination by other materials and does not change the response of typical test methods for which the material is intended.

Whenever possible, producers should undertake experimental verification (including measurement, expert inspection or qualitative testing as appropriate) to confirm the identity assigned using provenance.

**EXAMPLE** DNA extracted from a bacterial culture grown from a single bacterium, which in turn has been obtained from a bank of reference strain <sup>[43]</sup> could be certified based on provenance, subject to confirmatory checks for contamination.

### 9.7.2 Materials certified for identity based on measurements

#### 9.7.2.1 General

When characterizing the identity of a substance based on measurements, several aspects should be borne in mind.

a) Identity is usually not a measurement result, but a conclusion drawn based on measurement results from one or several methods. For example, chemical shifts and the heights of peaks in an NMR-spectrum, or a combination of colour, melting point, molar mass, etc. can inform an assignment of identity. While measurement uncertainties can be assigned to the individual measurement results, combining them to give any numerical indication of

(například pravděpodobnost, že přiřazená nejistota je správná) není jednoduché.

**PŘÍKLAD 1** Identifikace na základě sekvenční analýzy DNA ilustruje rozdíl mezi nejistotami v identitě a nejistotami přiřazenými experimentálním výsledkům. Daná DNA sekvence je výsledkem sekvenční analýzy a pravděpodobnost základního páru a jiných chyb při sekvenování, rovněž jako přítomnost mutačních rozdílů mezi jednotlivými molekulami DNA lze v principu odhadnout. Avšak identita biologických specií se může často stanovit se značnou spolehlivostí při relativně nízkém procentu homologie s referenční sekvencí. Jednotlivé sekvenční chyby tedy nemusí materiálně ovlivnit přiřazení identity.

- b) CRM certifikovaný na identitu je v praxi užitečný pouze tehdy, pokud pravděpodobnost chyby konečného úsudku je zanedbatelná.
- c) Nepatrná heterogenita a nestabilita materiálu nemusí měnit závěry o identitě. Hlavní podstatou hodnocení homogeneity a stability je aplikovatelnost daného materiálu, tedy zda stále dovolí jednoznačnou identifikaci.
- d) Různé látky mohou vykazovat stejné vlastnosti u vybraných identifikačních metod. Informace o zdroji surového materiálu a o krocích jeho zpracování před charakterizací je tudíž zásadní pro certifikaci na identitu.
- e) Jako u každého materiálu, plán projektu musí stanovit jasnou definici potřeb informací o identitě, která vychází ze zamýšleného použití materiálu.

**PŘÍKLAD 2** U DNA zamýšlené použití může vyžadovat pouze prohlášení o identitě specií, úplnou sekvenci, nebo dodatečnou informaci o stupni metylace.

**POZNÁMKA** Identita se někdy určuje na základě odborného posudku (například u azbestových vláken nebo druhů mikrobů). Ale toto posouzení se obvykle zakládá na pozorováních a porovnáních se specifikacemi. Odborný posudek vycházející z pozorování patří do rámce tohoto článku.

### 9.7.2.2 Specifikace

Testování identity materiálu představuje porovnání souboru výsledků měření na tomto materiálu se specifikacemi pro tyto výsledky měření (například rozmezí bodu tání, procento homologie s ohledem na DNA sekvenci).

**PŘÍKLAD** Organický polymerní materiál by se měl identifikovat pomocí porovnání s referenčním infračerveným spektrem (IČ) za pomoci těchto kritérií:

- frekvence všech píků se shodují s referenčním spektrem na  $3 \text{ cm}^{-1}$ ,
- relativní intenzity se shodují s referenčním spektrem na 5 % absorbance,

uncertainty in the identity (for example, a probability that the assigned identity is correct) is not straightforward.

**EXAMPLE 1** Identification using DNA sequencing illustrates the difference between uncertainties in identity and uncertainties associated with experimental results. The DNA sequence is a result of a sequence determination experiment, and the probability of base pair and other errors in the sequencing as well as the presence of mutation differences between different DNA molecules can in principle be estimated. However, identity of a biological species can often be established with considerable confidence at relatively low percentage of homology with a reference sequence. Individual sequencing errors therefore might not materially affect the assignment of identity.

- b) A CRM certified for identity is in practice only useful if the error probability on the conclusion is negligible.
- c) Slight heterogeneity and instability of the material may not change the conclusion of identity. The guiding principle for the assessment of homogeneity and stability is applicability of the material, i.e. whether it still allows unequivocal identification.
- d) Different substances can share the same properties for the identification methods chosen. Information on the source of the raw material and on the processing steps of the material to be characterized is therefore vital for the certification of identity.
- e) As with any material, the project planning shall establish a clear definition of the need for identity information based on the intended use of the material.

**EXAMPLE 2** For DNA, the intended use could require only a statement of species identity, a complete sequence, or additional information on the degree of methylation.

**NOTE** Identity is sometimes determined by expert judgment (e.g. for asbestos fibres or microbe species). However, this judgment is usually based on observations and comparison with specifications. Expert judgement based on observations falls within the scope of this clause.

### 9.7.2.2 Specification

Testing for identity of a material involves comparison of a set of measurement results on that material with specifications (for example, melting point range; percentage of homology with a reference DNA sequence) for these measurement results.

**EXAMPLE** An organic polymer material might be identified based on comparison with a reference infrared (IR) spectrum using the following criteria:

- all peak frequencies in the reference spectrum are matched within  $3 \text{ cm}^{-1}$ ;
- relative peak intensities match the reference spectrum within 5 % absorbance;

- žádné píky z referenčního spektra nechybí,
- všechny píky přítomné ve spektru kandidátského RM jsou též přítomny v referenčním spektru.

Jako zdroje specifikací mohou sloužit mezinárodně uznávaná kompendia (např. Lékopisy jako <sup>[44]</sup>) nebo další sbírky referenčních údajů (např. <sup>[45]</sup>). Změny takových informací nemusí mít výrobce RM pod kontrolou. Výrobci by měli tudíž jasně uvést použité specifikace pro hodnocení identity, ať již jako soubor hodnot nebo jako datovaný odkaz na certifikát u externí specifikace.

Při shromažďování specifikací by měli výrobci porovnat různé údaje z literatury a stanovit a zdokumentovat specifikace pro každou měřenou veličinu, které budou odrážet rozsahy a spolehlivost použitých informací. Přednost by se měla dávat údajům, které prošly recensním řízením.

### 9.7.2.3 Charakterizace kombinací metod

Tento přístup je zejména vhodný pro definované chemické látky s nízkou nebo střední molekulární hmotností.

Mělo by se vybrat několik metod, které budou prošetřovat různé vlastnosti kandidátského referenčního materiálu. Mezi často používané metody patří například stanovení bodu tání, molární hmotnosti, UV, IR, NMR a hmotnostní spektra. Spolu s informacemi o výchozí surovině a krocích jejího zpracování a odběrem vzorků a přepravou k výrobci RM by měl soubor takových metod postačovat nade vší pochybnost ke stanovení identity daného materiálu. Pokud existují podrobné publikované specifikace (např. kritéria Lékopisu pro identifikaci) může být výběr metod omezen na ty, které jsou uvedeny u těchto specifikací.

**POZNÁMKA 1** Podstata a počet požadovaných metod ke stanovení identity se mění s počtem potenciálně podobných produktů (například je více organických než anorganických látek) s informacemi o původu materiálů a krocích zpracovávání. Není tedy možno poskytnout žádná jasná pravidla pro výběr metod.

Všechny používané zkušební a měřicí postupy by měly být patřičně validovány a výsledky by měly splňovat požadavky na návaznost uvedené v odstavci 9.2. U všech postupů by se měly provést ověřovací experimenty za použití jiných referenčních materiálů, aby se prokázala správná kalibrace.

Výsledky každé z provedených zkoušek a měření by se měly porovnat se specifikacemi navrhované látky. Tam, kde jsou k dispozici, měly by se dodržovat publikované postupy pro taková porovnávání. Tam, kde takové předepsané postupy neexistují, výsledky měření by se neměly lišit v žádné ze specifikovaných hodnot, bere-li se v úvahu daná kombinovaná nejistota měření a specifikovaná hodnota.

- no peaks in the reference spectrum are absent;
- All peaks present in the candidate RM spectrum are present in the reference spectrum.

Sources of specifications can include internationally recognized compendia (e.g. Pharmacopeia sources such as <sup>[44]</sup>) and other collections of reference data (e.g. <sup>[45]</sup>). Such information can change outside the control of the RM producer. Producers should therefore clearly state the specifications used for the assignment of identity, either as a set of values or as a dated reference on the certificate to an external specification.

When compiling specifications, RM producers should compare various literature data, establish the range of reported values and establish and document specifications for each measurand reflecting the ranges and reliability of the information used. Preference should be given to reference data which have undergone peer review.

### 9.7.2.3 Characterization by a combination of methods

This approach is especially suitable for defined chemical substances of a small to medium molecular mass.

A number of methods should be chosen that probe different properties of the candidate reference material. Frequently used methods include, for example, determination of melting point, molar mass, UV, IR, NMR and mass spectra. Together with information on the raw material and its processing steps and the sampling and transport to the RM producer, the collection of methods should be sufficient to establish the identity of the material beyond any reasonable doubt. If detailed published specifications (e.g. Pharmacopoeial criteria for identification) exist, the choice of methods may be restricted to those listed in these specifications.

**NOTE 1** The nature and number of methods required to establish identity varies with the number of potentially similar products (e.g. there are more organic than inorganic substances) and the information on the origin and processing steps. Therefore, no clear rules can be given on the selection of methods.

All test and measurement procedures used should be properly validated and the results should fulfil the requirements for traceability laid out in 9.2. Verification measurements using other reference materials should be performed for all procedures used to demonstrate proper calibration.

The results of each of the tests and measurements made should be compared with the specification for the proposed substance. Published procedures for such comparisons should be followed where available. Where no such prescribed procedures exist, measurement results should not differ from any of the specified values when taking the combined uncertainty of measurement and specified value into account.

Jestliže výsledky souhlasí se specifikací, identita je stanovena se zanedbatelnou nejistotou.

**POZNÁMKA 2** Posouzení, zda nashromážděná měření a informace o provenienci budou postačovat ke stanovení identity nade vší pochybnost, je poněkud subjektivní. Výrobci RM se tak důrazně doporučuje, aby zavedli systém recenzních posuzování.

### 9.8 Přítomnost/nepřítomnost

Přítomnost/nepřítomnost je typem kvantitativního měření, které se vyhodnocuje kvalitativním způsobem. Výsledky překračující předem stanovenou mez se klasifikují jako „přítomen“, výsledky pod tímto prahem jako „nepřítomen“.

**POZNÁMKA 1** Mnohá měření se vyhodnocují jako přítomen/nepřítomen a nikdy se nekvantifikují. Avšak i u těchto metod existuje mezní hodnota odezvy, která se považuje za indikaci „přítomen“. Existence takové meze indikuje kvantitativní podstatu měření přes kvalitativní vyhodnocování.

Kvantitativní hodnocení takových měření je jedno řešení tohoto problému. Ačkoliv nejistota měření je obvykle vysoká (což je důvodem pro kvalitativní vyhodnocení), má tento přístup výhodu v tom, že je koncepčně jednoduchý. V nejjednodušším případě je charakterizace materiálu na nepřítomnost látky, pro kterou existuje kvantitativní metoda (např. kontaminant v potravinách). V tomto případě musí všechna měření daného materiálu poskytovat výsledky pod kritickou hodnotou a certifikovaná hodnota se uvádí jako „ $< L_d$ “, kde  $L_d$  je mez detekce.

**POZNÁMKA 2** Termín „mez detekce“ se používá ve smyslu definice IUPAC jako „nejmenší míra, kterou lze detekovat s přiměřenou jistotou u daného analytického postupu“<sup>[46]</sup>. Vztahuje se to k (pravé) koncentraci, kde existuje přiměřená jistota detekce látky, když je přítomna a odpovídá to CC $\beta$  (v rozhodnutí Komise 2002/657EC<sup>[47]</sup>) a „minimální detekovatelná hodnota redukované stavové proměnné“ (jak definuje ISO 11843-1<sup>[48]</sup>). U daného postupu tato mez závisí na počtu opakovaných měření.

Prohlášení o nejistotě by mělo obsahovat konfidenční úroveň pro jakoukoliv z udávaných horních koncentrací.

Pokud se používá několik metod a všechny výsledky souhlasí, může se použít mez detekce nejcitlivější metody jako certifikovaná hodnota.

**PŘÍKLAD** Tři různé metody poskytují výsledky uváděné jako  $< 2$  mg/kg;  $< 5$  mg/kg and  $< 4$  mg/kg. Všechny výsledky jsou v souladu, certifikovaná hodnota byla nastavena jako  $< 2$  mg/kg. Pro příklad viz odkaz<sup>[49]</sup>.

Použití různých postupů a/nebo laboratoří se doporučuje, aby se vyhnulo nesprávnému výkladu ztrát při analytickém postupu jako nepřítomnosti. Dále začlenění informací o zpracování daného materiálu se

If the results agree with the specification, identity is established with a negligible uncertainty.

**NOTE 2** A judgement on whether the accumulated measurement and provenance information is sufficient to establish identity beyond reasonable doubt, is somewhat subjective. RM producers are therefore strongly encouraged to establish a system of peer review.

### 9.8 Presence/absence

Presence/absence is an example of a quantitative measurement that is evaluated in a qualitative manner. Results above a predetermined threshold are classified as “presence”; results below are classified as “absence”.

**NOTE 1** Many measurements are evaluated as present/absent but are never quantified. Even for these methods however, there usually is a limit for the response to be regarded as indicating “presence”. Existence of such a limit indicates the quantitative nature of the measurement despite the qualitative evaluation.

Quantitative evaluation of the measurements is one solution to this problem. Although the measurement uncertainty is frequently high (which is the reason for the qualitative evaluation), this approach has the advantage of being conceptually simple. The simplest case is characterization of a material for the absence of a substance for which quantitative methods exist (e.g. a contaminant in a foodstuff). In this case, all measurements on the material have to give results below the critical value for declaring a substance present and the certified value is stated as “ $< L_d$ ” with  $L_d$  being the limit of detection.

**NOTE 2** The term “Limit of Detection” is used in its IUPAC definition as “smallest measure that can be detected with reasonable certainty for a given analytical procedure”<sup>[46]</sup>. This refers to the (true) concentration where one is reasonably certain to detect a substance if it is present and corresponds to CC $\beta$  (in European Commission Decision 2002/657EC<sup>[47]</sup>) and to the “minimum detectable value of the net state variable” (as defined in ISO 11843-1<sup>[48]</sup>). For a given procedure, this limit depends on the number of replicate measurements.

The uncertainty statement should state the confidence level for any upper limit given for the concentration.

If several methods are used, and the results all agree, the limit of detection of the most sensitive methods may be used as certified value.

**EXAMPLE** Three different methods give results stated as  $< 2$  mg/kg;  $< 5$  mg/kg and  $< 4$  mg/kg. As all results agree, the certified value is set as  $< 2$  mg/kg. For an example, see Reference<sup>[49]</sup>.

Use of different procedures and/or laboratories is advisable to avoid misinterpreting losses during the analytical procedure, as absence. Furthermore, inclusion of information on the processing of the

může využít k podpoře tvrzení, že určitá látka není přítomna.

Jestliže kvantifikace výsledků měření není možná, musí všechna měření poskytovat výsledek „nepřítomen“ k certifikaci materiálu na nepřítomnost určité látky a referenční hodnota v tomto případě se uvádí jako „nepřítomen“. Též by se měla uvést mez stanovitelnosti.

**POZNÁMKA 3** Při použití postupů detekční schopnosti podle ISO 11843-1,<sup>[48]</sup> „prohlášení méně než  $L_d$ “ má spolehlivost  $1-\beta$ , je-li pozorovaný výsledek pod kritickou hodnotou pro deklaraci přítomnosti látky a  $L_d$  se vypočítala tak, že pravděpodobnost výsledku pod kritickou hodnotou je  $\beta$ . Pokud je výsledek nad kritickou hodnotou, měla by se látka uvádět jako přítomna i když výsledek je pod  $L_d$ . Za těchto okolností je též technicky platné zachovat prohlášení, že látka je přítomna na úrovni nižší než  $L_d$ , ale na konfidenční úrovni, která je nižší než  $1-\beta$ .

### 9.9 Pořadové veličiny

Pořadové veličiny jsou veličiny, kdy výstup měření zařazuje výsledek do určité pořadové třídy. Mezi příklady patří Mohsova stupnice tvrdosti nebo podráždění kůže klasifikované jako žádná reakce /mírné zčervenání, významné podráždění/ prudká reakce. Tyto stupnice se obvykle definují ve vztahu ke konkrétní metodě klasifikace. Jedině možný přístup pro charakterizaci je tudíž charakterizace v několika laboratořích za použití stejné metody.

V mnoha případech bude CRM užitečný pouze tehdy, jestliže je zařazen jednoznačně do jediné třídy. Aby se toho dosáhlo, musí všechna technicky přijatelná měření u všech účastnících se laboratoří zařadit materiál do stejné třídy.

Jestliže se některé výsledky odlišují a odlišnost nelze vysvětlit technickými; chybami, nemůže se spolehlivě přiřadit žádná třída. Může však být užitečné poskytnout pro informaci medián a/nebo modus technicky platných výsledků

### 9.10 Kvalitativní parametry

Materiály se mohou charakterizovat na kvalitativní parametry, jako jsou barva, vůně, tvar atd. V některých případech se mohou tyto parametry kvantifikovat a jsou používány v praxi často v této formě. Příkladem jsou parametry tvarů částic nebo barva podle Hunterova systému.<sup>[50]</sup> To pak převádí problém na charakterizaci měřené veličiny definované postupem, jak se popisuje výše. Zvláště u barvy se může též zvažovat charakterizace pomocí spektra absorbance/reflektance.

### 9.11 Charakterizace necertifikovaných hodnot

Podle definice „referenčního materiálu“ a „certifikovaného referenčního materiálu“ musí být pouze certifikované hodnoty doprovázeny nejistotou měření a prohlášením o metrologické návaznosti.

material can be used to support the statement of absence of a certain substance.

If the quantification of the measurement results is not possible, all measurements have to provide the result “absent” to certify a material for the absence of a certain substance and the reference value in this case is stated as “absent”. Also, the limit of determination of the method should be given.

**NOTE 3** Using the capability of detection procedures of ISO 11843-1,<sup>[48]</sup> the statement “less than  $L_d$ ” has confidence  $1-\beta$  if the observed result is below the critical value for declaring the substance to be present and  $L_d$  has been calculated such that the probability of a result appearing below the critical value is  $\beta$ . If the result is above the critical value, the substance should normally be declared to be present even if the result is below  $L_d$ . Under that circumstance, however, it is also technically valid to retain the statement that the substance is present at a level below  $L_d$  but with a level of confidence that is less than  $1-\beta$ .

### 9.9 Ordinal quantities

Ordinal quantities are quantities where the outcome of the measurement puts the result in a certain ordered class. Examples are the Mohs' hardness scale or skin irritation classified as no response/moderate redness/significant irritation/severe reaction. These scales are often defined by reference to a particular method of classification. The only possible characterization approach in that case is therefore characterization by several laboratories using the same method.

In many cases, a CRM will only be useful if it is unequivocally put into one class. To achieve this, all technically accepted measurements by all participating laboratories shall put the material into the same class.

If some results deviate and the deviation cannot be explained by technical errors, no class can reliably be assigned. It can, however, be useful to give the median and/or the mode of the technically valid results as an information value.

### 9.10 Qualitative parameters

Materials can be characterized for qualitative parameters such as colour, odour, shape, etc. In some cases, these parameters can be quantified and are in practice often used in this quantified form. Examples are the shape parameters of particles or colour according to the Hunter system.<sup>[50]</sup> This transforms the problem to the characterization of a method-defined measurand as described above. Especially for colour, characterization of the absorbance/reflectance spectrum may also be considered.

### 9.11 Characterization of non-certified values

According to the definition of “reference material” and “certified reference material”, only certified values need to be accompanied by a measurement uncertainty and a statement of metrological traceability. Apart from the

Kromě certifikovaných hodnot se mohou uvádět necertifikované hodnoty (nazývané například „indikativní hodnoty“, „informační hodnoty“, „informativní hodnoty“ atd.), které se však nemohou používat jako reference v řetězci metrologické návaznosti.

Protože neexistuje požadavek na nejistotu a návaznost těchto hodnot (tj. neexistuje potřeba porovnávání) jakýkoliv přístup je přijatelný např. počínaje údaji z literatury o typických vlastnostech nebo příležitostných údajů z jediné laboratoře až k sloučeným datům z několika laboratoří. Čím více se bude podobat zvolený přístup k charakterizaci přístupu odpovídajícímu charakterizaci certifikovaných hodnot, tím spolehlivěji se bude taková necertifikovaná hodnota jevit.

Doporučuje se tedy poskytovat informace o původu necertifikovaných hodnot, aby uživatel mohl zhodnotit jejich spolehlivost. I když to není vyžadováno prohlášení o návaznosti a prohlášení o nejistotě zvýší užitečnost těchto hodnot.

## 10 Vyhodnocení nejistoty měření

### 10.1 Základ vyhodnocení nejistoty hodnoty vlastnosti CRM

Základními principy pro vyhodnocení nejistoty měření v tomto pokynu jsou obecné principy vydané v GUM.<sup>[51]</sup> GUM uvádí postup jak slučovat různé příspěvky, vyjádřené jako standardní nejistoty, do nejistoty. Shrnutí tohoto postupu uvádí příloha D.

Tento detailní postup je na místě, pokud může být popsán úplný model měření a všechny zdroje nejistoty jsou přidruženy jednotlivým ovlivňujícím veličinám daného modelu. V mnoha případech výroby referenčních materiálů však některé anebo všechny hlavní nejistoty nejsou spojeny s měřenými vstupními veličinami a je vhodné použít zjednodušený model, takový jaký se uveden v odstavci 10.2.

### 10.2 Základní model pro charakterizaci šarže

Hodnota certifikované vlastnosti v jednotlivé jednotce RM, když je dodána uživateli, může, v principu, být ovlivněna procesem charakterizace, reálnou variabilitou mezi jednotlivými jednotkami (heterogenitou), změnou v čase a změnami během přepravy a následného skladování. Model používaný pro vyhodnocení takové nejistoty u certifikované hodnoty musí zohledňovat všechny tyto vlivy tam, kde jsou významné. Vhodný jednoduchý model pro tento účel je následující:

$$x_{\text{CRM}} = x_{\text{char}} + \delta x_{\text{hom}} + \delta x_{\text{its}} \quad (19)$$

kde

$x_{\text{CRM}}$  značí hodnotu vlastnosti,

$x_{\text{char}}$  značí hodnotu vlastnosti získanou certifikací šarže nebo v případě charakterizace

certified values, non-certified values (named, for example “indicative values”, “information values”, “informative values”, etc.) may be assigned, which, however, cannot be used as a reference in a metrological traceability chain.

As there is no requirement for uncertainty and traceability (i.e. no requirement for comparability) of such values, any approach, ranging, e.g. from literature data about typical properties or circumstantial data from single laboratories to pooled data from several laboratories, is acceptable. The closer the chosen characterization approach resembles an approach appropriate for certified values, the more reliable this assigned non-certified value will be.

Therefore, it is recommended to give information on the origin of the non-certified value to allow users an assessment of its reliability. While not required, traceability statements and statements of uncertainties increase the usefulness of these values.

## 10 Evaluating measurement uncertainty

### 10.1 Basis for evaluating the uncertainty of a property value of a CRM

The basic principles for evaluation of measurement uncertainty in this Guide are the general principles set out in the GUM.<sup>[51]</sup> The GUM provides a procedure for combining different contributions to uncertainty, each expressed as standard uncertainties. A summary of the procedure is given in Annex D.

This detailed procedure is appropriate where a complete measurement model can be written and where all sources of uncertainty are associated with influence quantities in the model. In many cases in RM production, however, some or all of the major uncertainties are not associated with measured input quantities and a simplified model, such as that given in 10.2, is appropriate.

### 10.2 Basic model for a batch characterization

The value of a certified property in a single unit of an RM when delivered to the user can, in principle, be affected by the characterization process, by real variation between individual units (heterogeneity), change over time and changes during transportation and subsequent storage. The model used for evaluating the uncertainty on a certified property needs to allow for all of these effects where they are significant. A convenient simple model for this purpose is as follows:

$$x_{\text{CRM}} = x_{\text{char}} + \delta x_{\text{hom}} + \delta x_{\text{its}} \quad (19)$$

where

$x_{\text{CRM}}$  denotes the property value;

$x_{\text{char}}$  denotes the property value obtained from the characterization of the batch or, in the

jednotlivého artefaktu, hodnoty vlastnosti tohoto artefaktu,

$\delta x_{\text{hon}}$  značí chybový člen, který náleží heterogenitě a zohledňuje variabilitu mezi jednotkami spolu s možným potřebným příspěvkem na heterogenitu uvnitř jednotek,

$\delta x_{\text{Its}}$  značí chybový člen reprezentující vlivy stability během podmínek skladování.

Obvykle všechny studie homogenity a stability jsou navrženy takovým způsobem, že hodnoty těchto chybových členů se mohou předpokládat rovné nule, ale jejich nejistoty se nule rovnat nemusí.

Vzorec (19) představuje jednoduchý aditivní model měření, na který lze snadno aplikovat zásady GUM. Za předpokladu nezávislosti proměnných, se může nejistota spojená s hodnotou vlastnosti CRM vyjádřit jako

$$u_{\text{CRM}} = \sqrt{u_{\text{char}}^2 + u_{\text{hon}}^2 + u_{\text{Its}}^2} \quad (20)$$

pomocí zákona šíření nejistoty z GUM.<sup>[51]</sup> O vyhodnocení těchto složek nejistoty se pojednává v tomto pokynu v 7.11 (pro  $u_{\text{hon}}$ ) a 8.7 ( $u_{\text{Its}}$ ). Další návod poskytuje kapitola 9 (charakterizace) pro vyhodnocení nejistoty  $u_{\text{char}}$  vznikající při charakterizaci hodnoty vlastnosti.

**POZNÁMKA 1** Za určitých okolností může být žádoucí začlenit do vztahů (19) a (20) další členy. Například pokud nemůže být přeprava pod dostatečně dobrou kontrolou, aby se zaručila zanedbatelná změna, pak může být dodán do uvedeného modelu další vliv  $\delta x_{\text{trn}}$  a přidružená nejistota  $u_{\text{trn}}$ .

**POZNÁMKA 2** Kombinovaná standardní nejistota přidružená k hodnotě vlastnosti CRM se může vztáhnout k době platnosti materiálu (viz 8.6).

**POZNÁMKA 3** Někdy je hodnota vlastnosti funkcí času, jako u referenčních materiálů certifikovaných na radioaktivní izotopy. Model, použitý pro vyhodnocení nejistoty takové certifikované vlastnosti pro tento typ CRM by měl adekvátně počítat s časovou závislostí dané certifikované hodnoty.

### 10.3 Zdroje nejistoty

Nejistota charakterizace  $u_{\text{char}}$  by měla náležitě zohledňovat všechny významné zdroje nejistoty vyskytující se u měřicího postupu (postupů) použitého pro charakterizaci. Jak GUM<sup>[51]</sup> tak odkaz<sup>[52]</sup> vyjmenovávají běžné zdroje nejistot. Často bylo u měřicího postupu již provedeno hodnocení týkající se nejistoty měření a výsledné vyhodnocení nejistoty je možno použít pro vyhodnocení nejistoty u hodnoty vlastnosti CRM.

Modely nejistoty měřících metod často obsahují spojené složky, tj. složky nejistoty, které závisí na několika dalších. Tyto spojené (agregované) složky mohou vést ke kovariancím (viz 10.1), i když se to

case of a single artefact characterization, the property value obtained for this artefact;

$\delta x_{\text{hon}}$  denotes an error term due to heterogeneity which includes between-unit variation together with any necessary within-unit heterogeneity allowance;

$\delta x_{\text{Its}}$  denotes an error term representing the stability effects during storage conditions;

Usually, any homogeneity and stability studies are designed in such a way that the values of these error terms can be assumed to be zero, but their uncertainties might not be zero.

Formula (19) provides a simple additive 'measurement model' to which the GUM principles can be readily applied. Assuming independence of the variables, the uncertainty associated with a property value of a CRM can be expressed a

$$u_{\text{CRM}} = \sqrt{u_{\text{char}}^2 + u_{\text{hon}}^2 + u_{\text{Its}}^2} \quad (20)$$

using the law of propagation of uncertainty in GUM.<sup>[51]</sup> The evaluation of these uncertainty components is covered in this Guide in 7.11 (for  $u_{\text{hon}}$ ) and 8.7 ( $u_{\text{Its}}$ ). Additional guidance is given in Clause 9 (characterization) on the evaluation of the uncertainty  $u_{\text{char}}$  arising from the characterization of the property value.

**NOTE 1** In some circumstances, it can be desirable to include additional terms to Formulae (19) and (20). For example, where transportation cannot be sufficiently well controlled to guarantee negligible change, a further effect  $\delta x_{\text{trn}}$  and associated uncertainty  $u_{\text{trn}}$  can be added to the model given here.

**NOTE 2** The combined standard uncertainty associated with the property value of the CRM can be related to the period of validity of the material (see 8.6).

**NOTE 3** Sometimes, the property value is a function of time, such as for reference materials certified for radioactive isotopes. The model used for evaluating the uncertainty associated with the property value of this type of CRM should account accordingly for the time dependence of the certified value.

### 10.3 Uncertainty sources

The characterization uncertainty  $u_{\text{char}}$  should take due account of all the relevant uncertainty sources encountered in the measurement procedure(s) used for characterization. Both the GUM<sup>[51]</sup> and Reference<sup>[52]</sup> list common uncertainty sources. Often, the measurement procedures have already been evaluated in terms of measurement uncertainty, and the resulting uncertainty evaluation can be applied for the evaluation of the uncertainty of a property value of a CRM.

Often, uncertainty models of measurement methods contain aggregated components, i.e. uncertainty components which depend on several others. These aggregated components may lead to covariances (see



nezdá, pokud je metoda používána pro rutinní měření. Vyhodnocení kovariancí a korelací je rozhodné pro získání správného odhadu kombinované standardní nejistoty spojené s hodnotou vlastnosti CRM. Pro usnadnění procesu zjišťování kovariancí se doporučuje zdokumentovat, které složky nejistoty jsou obsažené ve spojených složkách nejistoty. Taková dokumentace pak umožní relativně rychlé určení možných zdrojů kovariancí a korelací. GUM (ISO/IEC Guide 98-3:2008, Annex F <sup>[51]</sup>) obsahuje určité další pokyny jak vyhodnocovat výsledné kovariance.

Jakákoliv změna konkrétního měřicího postupu by měla být doprovázena revizí modelu pro nejistotu.

#### 10.4 Intervaly pokrytí a koeficienty

Následně po výpočtu standardní nejistoty podle vzorce (20) je obvykle nutné vypočítat rozšířenou nejistotu  $U_{CRM}$ , aby se získal interval  $x_{CRM} \pm U_{CRM}$  zahrnující velký podíl hodnot, které mohou být důvodně přičítány vlastnosti, která je předmětem certifikace. Rozšířená nejistota se vypočítá za použití

$$U_{CRM} = k u_{CRM} \quad (20)$$

kde  $k$  je koeficient rozšíření.

Použitý koeficient rozšíření se určuje na základě předpokládané distribuční funkce a požadované pravděpodobnosti pokrytí. V mnoha případech lze předpokládat přibližně normální rozdělení, a pokud požadovaná pravděpodobnost pokrytí je rovna 95 %, vede to ke koeficientu rozšíření  $k = 2$ .

Tam, kde je počet efektivních stupňů volnosti spojených s  $u_{CRM}$  nízký (méně než 10), mělo by se používat Studentovo  $t$ -rozdělení namísto přiřazení koeficientu rozšíření rovnajícího se 2.

U případů, kde přiřazené rozdělení k hodnotě vlastnosti je považováno za asymetrické, jako je v případě výsledků zjišťování počtu sledujících Poissonovo rozdělení, měl by se uvádět interval pokrytí spíše než rozšířená nejistota s koeficientem rozšíření.

**POZNÁMKA** Intervaly pokrytí u asymetrických rozdělení bývají obvykle asymetrické.

10.1), even if these do not appear when the measurement method is used for a routine measurement. The evaluation of covariances and correlations is crucial to obtain a correct estimate of the combined standard uncertainty associated with the property value of a CRM. To facilitate the process of detecting covariances, it is recommended to document which uncertainty components are contained in the aggregated uncertainty components. This documentation allows relatively quick identification of possible sources of covariances and correlations. In the GUM (ISO/IEC Guide 98-3:2008, Annex F <sup>[51]</sup>), some further guidance is given on how to evaluate the resulting covariances.

Any change in a particular measurement procedure should be accompanied by a review of the uncertainty model.

#### 10.4 Coverage intervals and factors

Following calculation of the standard uncertainty  $u_{CRM}$  in Formula (20), it is usually necessary to calculate an expanded uncertainty  $U_{CRM}$  in order to provide an interval  $x_{CRM} \pm U_{CRM}$  including a large fraction of the values that could reasonably be attributed to the property being certified. The expanded uncertainty is calculated using

$$U_{CRM} = k u_{CRM} \quad (20)$$

where  $k$  is a coverage factor.

The coverage factor used is determined on the basis of the distribution function assumed and the coverage probability required. In many cases the distribution can be assumed to be approximately normal and the required coverage probability is 95 %; this leads to a coverage factor  $k = 2$ .

When the effective degrees of freedom associated with  $u_{CRM}$  is low (less than 10), the Student's  $t$ -distribution should be used instead of assigning a coverage factor of 2.

In cases where the assigned distribution of the property value is considered to be asymmetric, such as in the case of the result of a count following the Poisson distribution, a coverage interval should be stated rather than the expanded uncertainty and a coverage factor.

**NOTE** Coverage intervals for asymmetric distributions will usually be asymmetric.

## Příloha A (normativní)

### Návrh a vyhodnocení studií pro charakterizaci na metodě nezávislé měřené veličiny pomocí dvou nebo více metod s prokazatelnou přesností v jedné nebo více kompetentních laboratořích

#### A.1 Návrh studie

##### A.1.1 Výběr metod

Při výběru metod by měla být vzata v úvahu variabilita následujících aspektů:

- příprava vzorku, např. drcení/mletí, extrakční techniky, rozpouštědla, vyluhovací prostředí, objemy/podmínky, čisticí kroky;
- zavádění vzorku a/ nebo jeho separace, např. pomocí LC nebo GC;
- kvantifikační principy, např. molekulová nebo atomová absorpce, hmotnostní spektrometrie, plamenová ionizace, fluorescence;
- kalibrační postupy, pokud jeden postup nevykazuje jasné výhody v důsledku jeho metrologické exaktnosti nebo v důsledku dosahování nízkých nejistot měření.

V mnoha případech, variabilita všech aspektů není možná. V těchto případech by měla být vyhledávána jejich maximální možná variabilita. Například, jestliže je plynová chromatografie jedinou možnou separační technikou, potom by studie měla usilovat o zahrnutí různých vstřikovacích technik, různých kolon a teplotních programů a kvantifikaci různými detektory.

Výrobce by měl požadovat, aby všechny metody použité v dané akci byly řádně validovány a aby přidružená nejistota měření mohla být srozumitelně doložena.

Aby bylo možné povolit laboratořím svobodný výběr metod při současném zajištění nezbytné jejich variability, výrobce by měl ještě před zahájením studie získat informace o metodách používaných jednotlivými participanty tak, aby získal nezbytnou řadu metod. Pokud je to žádoucí, mělo by být provedeno cílené vyhledávání laboratořích nabízejících specifické metody, aby se předešlo situaci, že přijaté výsledky budou získány pouze jednou metodou.

##### A.1.2 Volba kalibračních standardů

Důležitým rozhodnutím je, zda by všechny laboratoře měly použít stejný kalibrátor nebo zdali by měl být laboratořím ponechán svobodný výběr kalibrátoru. Použití jediného kalibrátoru redukuje variabilitu způsobenou různými kalibrátory od různých dodavatelů. Na druhé straně, jakékoliv vychýlení u toho jediného kalibrátoru bude přeneseno do stejného vychýlení u certifikovaných hodnot. Tudíž použití jediného kalibrátoru vyžaduje velmi pečlivou charakterizaci tohoto kalibrátoru, vyžadující v mnoha případech nezávislé potvrzení jeho čistoty nebo složení. Ponechat laboratořím svobodnou volbu kalibrátoru je logicky snazší, nevyžaduje confirmaci nezávislými testy a umožňuje laboratořím aplikovat jejich pracovní postupy nezměněny, znamená to ale také, že do studie by měly být zahrnuty nějaké cesty k ověření kalibrace.

Obecně platí:

- pro dobře zavedená měření, kde zkušenosti ukazují, že kvalita dostupných kalibračních standardů je dostatečná, se obvykle preferuje svobodná volba standardů jednotlivými laboratořemi;
- v případech, kdy existuje značná pochybnost o kvalitě standardů na trhu, jsou často oprávněné snahy po charakterizaci běžně používaného standardu.

V některých případech, různí výrobci komerčních standardů získávají jejich čistý materiál od stejné společnosti. Je tudíž užitečné zjistit, zdali se zdroje původu liší.

##### A.1.3 Výběr laboratořích

Laboratoře by měly být vybírány na základě jejich prokazatelné kompetence. Vhodným důkazem prokazování kompetence mohou být:

- výsledky z předchozích testů způsobilosti;
- výsledky získané s nezávislými CRM (pravděpodobně distribuované jako materiály pro řízení kvality společně s kandidátským CRM);
- údaje z validace metod;
- úplná a věrohodná bilance nejistot;

- předchozí účast v certifikačních studiích jiných RM pro stejnou měřenou veličinu a
- osvědčení třetí stranou v souladu s ISO/IEC 17025 [53] nebo s jinými relevantními normami pro stanovení předmětné měřené veličiny.

Výkonnost externě posuzovaných laboratoří se může lišit stejným způsobem jako u ostatních kompetentních laboratoří. Je tudíž prozíravé získat kromě osvědčení třetí stranou v souladu s ISO/IEC 17025 také informace o výkonnosti laboratoře.

Mělo by být doloženo, že měření v každé laboratoři bylo provedeno v souladu s ISO/IEC 17025. Zejména by měla být naplněna ustanovení ISO/IEC 17025 týkající se kompetence personálu, kalibrace přístrojů a autorizace metod a výsledků (úřední vydání pro použití k souladu s ISO/IEC 17025 8:2005,4.3.2.1. [53]).

V případě absence nezávislého osvědčení by měla být získána informace o rozsahu systému kvality laboratoře.

**POZNÁMKA 1** Potenciální (eventuální) mezery ve splnění požadavků určité normy mohou být někdy vyplněny výrobcem RM (např. výrobce RM může archivovat původní údaje z laboratoří, pokud nemají vlastní archivační systém).

**POZNÁMKA 2** Získání kalibračních certifikátů pro každý přístroj je v mnoha případech nepraktické. Shoda výsledků získaných pomocí materiálů pro řízení kvality může být použita jako důkaz dostatečné kalibrace všech relevantních vstupních faktorů.

#### A.1.4 Počet nezávislých souborů dat

Počet zúčastněných laboratoří je méně důležitý než počet nezávislých souborů dat. Jedna laboratoř může být schopna dodat několik souborů dat, všechny získané nezávislými metodami. Zbývající část této diskuse se zaměří na soubory dat, bez ohledu na to, zdali každý soubor dat byl poskytnut různou laboratoří, některé laboratoře poskytly více než jeden soubor dat nebo všechny soubory dat byly vyprodukovány ve stejné laboratoři.

Je těžké dosáhnout úplné nezávislosti výsledků, pokud jsou měření provedena v jedné laboratoři. Výrobce by měl kriticky zrevidovat variabilitu všech kritických kroků uvedených v odstavci A.1.1., aby zjistil, zdali je dosaženo dostatečné variability metod a tím potvrdit nepřítomnost vychýlení v každém jednotlivém kroku. To zahrnuje i zjištění, zdali stejná problematická chemikálie, přístroj, kalibrátor atd. nebyla použita v různých metodách. Organizátor by měl na začátku studie rovněž laboratořím připomenout, aby necenzurovaly data, tj. aby po vzájemné kontrole výsledků mezi jednotlivými metodami nepotlačovaly/neměnily (bez předchozího upozornění) data získaná jednou metodou.

Výrobce by měl stanovit minimální doložený počet technicky platných výsledků, z nichž bude rozhodovat o přidělení hodnoty veličiny. Počet souborů dat by měl být dostatečně velký, aby mohl poskytnout přiměřeně nízkou nejistotu odhadované hodnoty a přitom umožnil existenci chyby při prezentování výsledků, vyřazení výsledků z technických důvodů a zamýšlené statistické vyhodnocení.

Pro mezilaboratorní studie se doporučuje naplánovat získání pěti nebo více nezávislých souborů technicky platných výsledků, které by měly sloužit jako základ pro výpočet certifikované hodnoty.

Následující úvahy mohou změnit požadovaný počet souborů dat tak, aby se dosáhlo požadované nejistoty certifikovaných hodnot.

- **Požadovaná nejistota:** Nejistota přiřazených hodnot měřené veličiny se obvykle snižuje s počtem technicky validovaných souborů měření, což vyžaduje větší soubory dat k získání nižších nejistot.
- **Technická obtížnost:** Čím méně je měření zavedeno nebo čím více je technicky náročné, tím větší se dá očekávat variabilita mezi soubory dat.
- **Pravděpodobnost technicky neplatných výsledků:** I zkušené laboratoře mohou dodat technicky neplatné výsledky, které nemohou být použity pro certifikaci. Takové výsledky jsou více pravděpodobné pro neznámé materiály, nové nebo upravené postupy měření, nová nadějná měření nebo neobvyklé požadavky na uvádění dat. Počet účastníků nebo (pro studie prováděné v jediné laboratoři) počet nezávislých měření by se měl zvýšit všude tam, kde hrozí vyšší riziko technických chyb.
- **Poměr reprodukovatelnost/opakovatelnost:** Tam, kde variabilita mezi soubory dat je hlavním zdrojem variability, měla by být dána přednost spíše většímu počtu souborů dat, než většímu počtu opakovaných měření v každém souboru dat.
- **Statistické vyhodnocení:** Různé způsoby vyhodnocení dat mohou vyžadovat větší množství pozorování, aby byla zajištěna dobrá numerická stabilita a/nebo dosažena dostatečně nízká nejistota.
- **Validace výsledků pomocí CRM:** Výsledky stabilně vysoké kvality, získané pomocí jednoho nebo více CRM, používaných pro zajištění kvality v každé laboratoři, slouží k prokázání platnosti všech souborů dat a poskytují informaci o vychýlení mezi laboratořemi a metodami. Tyto informace dovolují aplikovat Bayesovské modely odhadu nejistoty, pro které stačí menší počet souborů dat ve srovnání s vyloženě četnostním přístupem.

### A.1.5 Počet jednotek a opakovaných stanovení

Počet jednotek kandidátského CRM zaslaný každému účastníkovi, počet opakovaných stanovení provedených každým účastníkem a podmínky těchto stanovení jsou stanoveny praktickými důvody, stejně tak jako záměrem o vyhodnocování dat.

- V případě, že se spotřebuje celá jednotka, nebo jestliže se výsledky průměrují ze všech jednotek, tak se částečně kompenzuje heterogenita mezi jednotkami. Jestliže je heterogenita příliš vysoká, upřednostňuje se jedno měření v několika různých jednotkách před několika opakovanými měřeními v jednotce. Jestliže nehrozí kontaminace, spotřeba celé jednotky nebo heterogenita, zaslání jedné jediné jednotky je postačující.
- V případě absence spolehlivého odhadu nejistot účastníky, požadavek provedení měření za podmínek mezilehlé preciznosti může poskytnout informaci o spolehlivosti výsledků účastníka nebo poskytovat kontrolu spolehlivosti odhadů nejistot jednotlivými účastníky.

### A.1.6 Materiály pro řízení kvality

I když studie zahrnuje pouze laboratoře s prokazatelnou kompetencí, přidání dalších vzorků pro řízení jakosti bylo shledáno velmi užitečným. Výsledky těchto vzorků mohou identifikovat technické problémy a napomoci technickému vyhodnocení.

- RM, zejména přírodní matricové RM a materiály pro řízení kvality (QCM) smějí být použity pro prokazování platnosti výsledků měření pouze tehdy, pokud jsou měřeny společně s neznámými vzorky, které mají být charakterizovány.
- Spikované materiály, spikované slepé pokusy atd. mohou být použity ke kontrole částí měřicího postupu nebo mohou být nápomocny při přidělování hodnot vlastností materiálu.
- Blanky matricových materiálů, blanky extraktů atd. mohou být použity k důkazu, že měřicí metoda poskytuje výsledky, které se významně neliší od nuly, když předmět zájmu není přítomen (jak je často prováděno při měření složení), nebo pro stanovení korekce nebo korekčního faktoru, společně s nejistotou korekčního faktoru.

POZNÁMKA Někdy se certifikované referenční materiály, používané ke kontrole kvality při mezilaboratorních studiích, dodávají bez originálních štítků, aby se předešlo jejich identifikaci. Jelikož je však množství spolehlivých certifikovaných referenčních materiálů omezené, zkušené laboratoře rozpoznají materiál podle jeho vzhledu a/nebo hodnot.

### A.1.7 Uvádění výsledků

Použití předem připravených formulářů pro zprávy může být užitečné, jelikož zpráva je pak již strukturována a (pokud je zaslána elektronicky) umožňuje kopírování výsledků, což značně snižuje možnost chyb při přepisování a shromažďování výsledků na straně výrobce RM. Nevýhodou předformátovaných formulářů je, že často nutí laboratoře opustit jejich prezentační zvyklosti, což může mít za následek chyby v přepisování dat. Jestliže jsou zprávy zasílány elektronicky, musí se trvat na dodržování požadavků Pokynu ISO 34 (2) na integritu elektronických záznamů, zejména pokud se jedná o prezentování výsledků testu.

POZNÁMKA: Když jsou zprávy předávány ve formátu tabulkového editoru, nežádoucím změnám lze často zabránit použitím „zamykacích“ nebo „ochranných“ prostředků zabudovaných do příslušného softwaru. Laboratoře by měly být požádány, aby uváděly jednotlivé výsledky (nejenom průměry všech vzorků), bez ohledu na to, zdali je nebo není uveden údaj o nejistotě, i když uvedení průměru a rozšířené nejistoty s příslušným koeficientem rozšíření by měla být dostačující.

Tam, kde existuje možnost korekce známého procedurálního vychýlení, jako je výtěžnost extrakce, výrobce by měl jasně stanovit, zdali mají být výsledky korigovány nebo ne. Výrobce by měl rovněž po účastnících požadovat, aby prezentovali kontroly vychýlení (např. pomocí přídatku) a použít je pro korekci výsledků je-li zjištěno vychýlení. Tam kde je korekce provedena účastníkem, jakákoliv prezentovaná nejistota by měla zahrnovat příspěvek nejistoty spojený s korekcí.

Účastníci by měli být instruováni jakým způsobem uvádět výsledky blízké mezi detekce, jestliže takové výsledky mohou nastat. Výsledky prezentované jako „méně než“ činí statistické vyhodnocení obtížnější. Na druhé straně, uvádění výsledků v okolí mezí detekce je v rozporu s postupy řízení kvality v mnoha laboratořích. Tam, kde jsou výsledky v okolí mezí detekce pravděpodobné, měli by výrobci buď požádat laboratoře, aby prezentovaly pozorované výsledky (včetně záporných) – místo nebo navíc jejich obvyklého prezentování, nebo aby přijaly statistické postupy, které umožňují prezentovat takové výsledky jako „méně než“, nebo výsledky omezit pouze na kladné hodnoty.

Doporučuje se, aby popis použitého měřicího postupu byl prezentován dostatečně podrobně s tím, aby umožnil pochopit části postupu měření (např. v chemické analýze loužení/extrakci vzorku a separaci jednotlivých analytů, čištění, kvantifikaci). Tam, kde je to na místě, účastníci by měli být požádáni, aby uvedli literární reference.

### A.1.8 Instrukce pro účastníky

Pokyny účastníkům by měly obsahovat:

- a) jasný záměr cíle studie;
- b) pokyny zdržet se porovnávání výsledků s ostatními účastníky, zahrnující důvody odrazující je od nekalých dohod (tj. vzájemnou výměnu informací);
- c) počet testovacích jednotek;
- d) počet opakovaných stanovení, které mají být provedeny;
- e) jaké měřicí metody mají být použity;
- f) velikost zkoušeného vzorku;
- g) požadavky týkající se kvality a návaznosti výsledků měření;
- h) časový plán (distribuce vzorků, dodání výsledků);
- i) způsob odeslání;
- j) pokyny pro mezilehlé skladování vzorků;
- k) specifické pokyny pro zpracování vzorku, připadá-li to v úvahu;
- l) popis mechanismů pro ověřování předpokladů týkajících se dat, a ...
- m) pokud je to žádoucí, prohlášení, že výsledky budou rozšiřovány pouze anonymním způsobem (tj. nebude žádná vazba mezi určitou laboratorí a výsledky) nebo že cíl měření je pokládán za splněný, i když z technických důvodů nejsou použity výsledky určité metody.

Schůzka se zainteresovanými laboratořemi /skupinami (před distribucí vzorků a provedením měření) může pomoci všem zúčastněným stranám sjednotit všechny akce prováděné během společné studie a diskutovat možné problémy a/nebo úskalí.

Organizátor studie by měl vzít v úvahu, že diskutování výsledků různých metod mezi pracovníky laboratoře je zcela běžné a může být užitečným krokem k odrazení dohod uvnitř laboratoře a mezi laboratořemi. Například, různé označení materiálů pro každého účastníka může učinit porovnání výsledků mezi účastníky těžší.

## A.2 Vyhodnocování

### A.2.1 Počáteční screening

Výrobce může kontaktovat účastníky, aby asistovali při zkoumání anomálií v jakékoliv fázi procesu vyhodnocování. Jestliže je účastník kontaktován, doporučuje se, že by počáteční kontakt neměl specifikovat povahu anomálie (např. směr odchylky výsledků), účastník by měl být spíše vyzván, aby zjistil a oznámil jakékoliv nalezené chyby.

V případě, že jedna laboratoř dodala soubory dat z více než jedné metody měření, počáteční prohlídka by měla posuzovat soubory dat individuálně (tak jakoby byly nezávisle prezentovány různými laboratořemi). Nicméně, výrobce RM by měl vzít v úvahu všechny laboratořemi dodané soubory dat při formulování závěrů, které soubory dat si ponechat v případě, že jsou nalezeny nějaké anomálie.

### A.2.2 Technické posouzení

Další technické zhodnocení za účelem identifikování potenciálních problémů může zahrnovat (ale nemělo by být omezeno k) vytváření skupin výsledků podle jednotlivých technik (měřících metod a principů metod, metod předúpravy vzorku, atd.) nebo použití různých kalibračních technik. Navíc, porovnání rozšířených nejistot s intervalem spolehlivosti průměru dodaných výsledků poskytuje svědectví, zdali jsou uvedené nejistoty realistické, protože rozšířená nejistota by měla být alespoň tak velká, jako interval spolehlivosti.

Technicky neplatný výsledek není nezbytně odlehlou hodnotou, ani každá odlehlá hodnota není nezbytně technicky neplatná. Výsledek může dobře zapadnout do intervalu platných výsledků, i když je zcela zřejmé, že podmínky, za kterých byl výsledek získán, nebyly v pořádku. Technicky neplatné výsledky by měly být vyjmuty z přijatých souborů dat, nebo pokud je to možné nebo nutné opraveny opakovaným měřením. Naopak výsledek, který se významně liší od všech ostatních výsledků, může být jediným technicky platným výsledkem z celého souboru dat.

### A.2.3 Statistické vyhodnocení

#### A.2.3.1 Obecné zásady pro statistické vyhodnocování

Charakterizace pomocí společné studie nebo mnoha měřícími postupy cílí na náhodnost vychýlení mezi soubory dat. Statistické vyhodnocení obecně předpokládá, že pravá hodnota měřené veličiny odpovídá pravé hodnotě parametru základního souboru, obvykle jeho výběrovému průměru.

Různé postupy odhadu určité měřené veličiny – zdali je měřená veličina definována postupem či nikoliv – mohou být systematicky vychýleny stejně tak jako specifické vychýlení souboru dat laboratoře. Možnost rozdílů mezi metodami by se měla obvykle zvažovat při vyhodnocování nejistoty měření.

### A.2.3.2 Rozdělení

Nalezení vhodných odhadů očekávané hodnoty je úzce svázáno s (buď předpokládaným, nebo stanoveným) základem rozdělení hodnot. Stanovení hustoty pravděpodobnosti ze souboru dat vyžaduje mnohem více dat, než je prakticky možné získat, zatímco ověření konzistence s předpokládaným rozdělením a výpočet parametrů z tohoto rozdělení (např. průměru normálního rozdělení) je možné z počtu souboru dat obvykle získaným z charakterizačních studií. Výrobce musí tudíž ověřit, zdali existuje nějaký důkaz o odchylce od předpokládaného rozdělení užívajíc např. vizuální metody (histogramy, Kernelovy diagramy hustoty, a normální pravděpodobnost nebo obecněji Q-Q diagramy) nebo statistické kontroly odklonu od určitých rozdělení, zahrnující testy normality nebo stanovení zešikmení a špičatosti. Přibližně normální rozdělení souborů dat je často pozorováno u výsledků dostatečně nad mezí stanovitelnosti, ostatní rozdělení zahrnují Poissonovo rozdělení (např. pro počet mikrobů) nebo Weibullovo rozdělení (např. mechanické poruchy v keramice). Zvolené rozdělení musí být ve shodě s prezentovanými daty stejně tak jako s historickými a teoretickými poznatky měření, o které se jedná. Jestliže se tyto významně liší, žádná hodnota nesmí být přiřazena, pokud nebudou udány technické důvody pro neočekávané rozdělení.

**PŘÍKLAD** Teoretické úvahy právě tak jako množství mezilaboratorních porovnání ukazují, že výsledky stanovení stopových prvků v půdě vykazují normální rozdělení (pokud není jejich obsah blízký mezí stanovitelnosti). Odchylna pozorovaných dat od normálního rozdělení (např. chvostování) ukazuje na technické problémy. Protože není zcela jasné, zdali problém leží ve většině nebo ve chvostech, žádnou hodnotu nejde přiřadit.

V některých případech výsledky mohou být upraveny tak, že se stanou normálně rozdělenými. Některé běžně používané transformace jsou logaritmování, aplikace druhých odmocnin a exponenciální formy. Pro takové transformace by však mělo existovat určité opodstatnění, protože odchylky výsledků od očekávaného rozdělení mohou indikovat technické problémy.

### A.2.3.3 Odlehlé výsledky

Odlehlé výsledky mohou nastat na všech úrovních společné studie: při jednotlivých pozorováních, v podskupinách (např. skupiny v jednotlivých lahvíčkách), nebo mohou být odlehlé „výsledky“ celých metod/laboratoří. K identifikaci odlehlých výsledků mohou být použity vhodné testy odlehlosti od průměru z rozptylu a rovněž Mandelova H a K statistika [24].

Odlehlá pozorování nebo hodnoty průměru by neměly být vyloučeny pouze na základě statistických testů odlehlosti, ale měly by být vyloučeny, jsou-li proto rovněž technické důvody.

**POZNÁMKA 1** Technické důvody zahrnují nedostatečnou kalibraci, nepřiměřené postupy měření, použití nepřiměřených činidel, chyby při posuzování interferencí, odchylku od certifikované hodnoty nezávislého materiálu pro kontrolu kvality, atd.

**POZNÁMKA 2** Testy odlehlosti obvykle ignorují nejistoty měření. Odlehlý soubor dat může být v souladu s ostatními daty, pokud se vezmou v úvahu odpovídající nejistoty.

**POZNÁMKA 3** Jednotlivá data s vysokými nejistotami mohou být vyřazena na technickém základě, i když souhlasí s ostatními výsledky v rámci jejich nejistoty a jestliže nejistota je tak vysoká, že to může indikovat technický problém, nebo mohou být ponechána, protože existuje shoda mezi výsledky.

**POZNÁMKA 4** Soubory dat, které vykazují vysokou variabilitu odlehlosti, mohou ukazovat na nedostatečnou opakovatelnost metody nebo nedostatečnou mezilehlou preciznost, což opravňuje jejich vyloučení z technických důvodů.

**POZNÁMKA 5** Různé metody se mohou lišit v preciznosti, takže může být důležité ponechat soubory dat z relativně neprecizní metody, aby zůstal zachován reprezentativní soubor metod.

Speciálním případem odlehlého rozptylu je soubor výsledků mající nulový rozptyl. Laboratoře jsou často omezeny v počtu uváděných desetinných míst. Takové omezení může být v rozporu s účelem projektu. Je žádoucí, aby byl tento problém předvídan při zadávání instrukcí pro prezentaci. Alternativně, laboratoř může poskytnout na požádání dodatečné číslice.

### A.2.3.4 Robustní statistika

Robustní statistika nabízí širokou škálu metod, které výslovně povolují přítomnost odlehlých hodnot v jinak přibližně normálně rozděleném souboru dat. Obecně, robustní metody přiřazují váhy, které se snižují se vzdáleností od hlavního výskytu dat. Robustní metody jsou určeny pro zjišťování průměru a směrodatné odchylky

v souborech dat kontaminovaných jednoduše odlehlými výsledky, stejně jako pro další parametry, které jsou předmětem zájmu. Robustní metody jsou také určeny pro další situace zahrnující např. soubory presentovaných hodnot se značně odlišnými nejistotami nebo pro analýzu dat složených z mnohonásobně opakovaných měření v každém souboru dat.

Robustní odhady vykazují značnou odolnost k vlivu extrémně odlehlých hodnot. Pro symetrické, „velmi chvostující“ rozdělení (které zahrnuje velký podíl hodnot vzdálených od průměru než by se očekávalo od normálního rozdělení), robustní statistika typicky vyazuje nevyčýlené odhady průměru s menším rozptylem než je tomu za stejných podmínek u jednoduchého aritmetického průměru. Pro data s extrémně odlehlými výsledky může být robustní statistika značným zlepšením ve srovnání s průměrem. Krátké shrnutí užitečných robustních statistických metod a podmínek pro jejich použití při charakterizaci RM a odkazy na podrobnější informace je uveden v kapitole B.5.

**POZNÁMKA** Odolnost robustní statistiky vůči extrémním hodnotám ji dělá kromě této užitečnou při identifikaci odlehlých výsledků ve velkých souborech dat obsahujících extrémní hodnoty.

### A.2.3.5 Shlukování

Jestliže výsledky tvoří dva nebo více zřetelných shluků (klastřů), měly by být vzaty v úvahu následující možnosti.

- Jestliže existuje korelace těchto shluků s měřicími metodami nebo postupy (nebo jestliže pouze jedna odchýlená hodnota pochází z odlišného principu metody) a jestliže rozdíl mezi průměry těchto shluků je statisticky významný a jestliže rozdíly mezi průměry těchto shluků jsou příliš velké na to, aby umožnily získání dostatečně malé nejistoty pro zamýšlené použití materiálu, pak nejde přiřadit jedinou hodnotu vlastnosti. V tomto případě je k řešení problému nezbytné vylepšit postupy měřicí metody; jinak by mohlo být ke každému postupu přiřazeno několik procedurálně definovaných hodnot vlastností.
- Jestliže existuje korelace těchto shluků s měřicími metodami nebo postupy, avšak rozdíl mezi průměry těchto shluků je relativně malý, jediná hodnota může být přiřazena. Dodatečný člen nejistoty, odpovídající rozdílu mezi metodami by měl být přidán k nejistotě charakterizace. Existuje několik přístupů k řešení tohoto problému popsanych v literatuře. [54] [55].
- Jestliže neexistuje korelace těchto shluků s měřicími metodami nebo postupy a jestliže rozdíl mezi těmito shluky je značný, nelze přiřadit žádnou hodnotu. Větší množství (zásoba) výsledků je nezbytné pro překonání relativně špatného souladu dostupných měřících metod a postupů.

## A.3 Použití společných studií pro další účely

**A.3.1** ISO 5725-3 [14] uvádí řadu účelů pro mezilaboratorní porovnání, zahrnující přiřazení hodnot referenčním materiálům, vyhodnocení práce laboratoří a vyhodnocení funkčních charakteristik metody. Všeobecně řečeno, určitá studie je nejlépe použitelná pouze pro jeden z těchto účelů, pokud je to provedeno jinak, může to znamenat kompromis u jednoho cíle ve prospěch jiného, nebo učinit účel studie pro účastníky zmatečným.

**A.3.2** Hlavním důvodem pro kombinování charakterizačních studií pro referenční materiály s ostatními studii jako jsou studie výkonosti metod a zkoušení způsobilosti (PT) je evidentně snížení nákladů.

**A.3.3** Potenciální neslučitelnosti nebo konflikty způsobené spojením PT nebo studií výkonosti metody s charakterizačními studii jsou následující:

- PT studie jsou zaměřeny na posuzování kompetence laboratoře, zatímco účast v charakterizační studii je omezena pouze na laboratoře s prokazatelnou kompetencí. Kombinace PT studie s charakterizační studii by tudíž mohla předpokládat právě skutečnost prokazatelné kompetence pro účely charakterizace, který by měl být posouzen v části PT.
- Požadavky na zkoušení způsobilosti běžně nevyžadují rozsáhlé opakování měření a zřídka kdy dovolují požadovat stejný počet opakovaných měření všemi účastníky.
- Výsledky PT jsou obvykle zpracovávány v utajení a tak nemůže být zaručen přístup k detailům experimentálních metod.
- Za účast v PT studii účastníci často platí a tak motivace účastníků může být někdy popoháněna komercí. Účastníci si proto často nepřejí, aby se šlo na detailním uspořádání studie, na požadavcích zajištění návaznosti a kontroly kvality, kalibraci, stanovení nejistot atd., které jsou vyžadovány u charakterizační studie.
- Účast v PT studii je požadavkem pro akreditaci a výsledky se používají pro prokazování kompetence potenciálním zákazníkům. Účastníci si proto mohou méně přát otevřeně diskutovat a připustit technické problémy tak jak je vyžadováno u charakterizační studie při technickém vyhodnocování výsledků.
- PT studie obvykle posuzují výkonost laboratoře ve srovnání s ostatním i laboratořemi nebo s externě nastavenými kritérii, zatímco charakterizační studie se snaží získat dobrý odhad právě hodnoty a její nejistoty.

To vede k rozdílným přístupům při vyhodnocování, zejména v zacházení s extrémními hodnotami/odlehlymi hodnotami.

- g) PT studie jsou otevřené pro každou laboratoř, která si přeje se zúčastnit. Výrobci mají proto často malou kontrolu nad metodami, které účastníci používají, což může vést k tomu, že dostanou výsledky pouze z jedné metody, i když existují i další metody.
- h) Zatímco zpětná vazba na výsledky za účelem zjištění anomálií nemůže nastat dříve než je PT studie uzavřena, u charakterizační studie je tato informace vyžadována pro sledování zřejmých anomálií výsledků
- i) Studie výkonosti metody vyhodnocují výkonost metody okamžitě po jejím vývoji, takže existuje málo předchozích informací o její typické výkonosti
- j) Vyhodnocení PT studie se často velmi mnoho soustředí na statistiku, zatímco charakterizační studie klade větší důraz na technické posouzení.
- k) Studie výkonosti metody většinou používají pouze jeden měřicí postup a nejsou tudíž vhodné pro přiřazování hodnot analytům nezávislých na metodě.

**A.3.4** V důsledku těchto potenciálních rozporů, dohodnuté hodnoty z PT programů by neměly být použity pro certifikaci referenčního materiálu bez kritického hodnocení dat. Data získaná od účastníků PT programů by však mohla být použita pro certifikaci referenčního materiálu za předpokladu, že jsou splněny všechny níže uvedené požadavky:

- a) rozhodnutí ještě před začátkem studie, která podskupina dat ze specifikovaných laboratoří bude použita pro přiřazení hodnoty;
- b) prokázání kompetence těchto laboratoří nezávisle na této studii, např. z výsledků v předchozích kolech tohoto programu nebo jinými prostředky;
- c) organizování studie pro tyto vybrané laboratoře podle kritérií popsanych v A.1.3;
- d) vyhodnocení výsledků těchto laboratoří podle kritérií uvedených v kapitole A.2.
- e) oznámení navrženým účastníkům, že jejich výsledky mohou být použity v certifikační studii a získání svolení k přístupu k technickým detailům.



## Annex A (normative)

### 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

#### A.1 Study design

##### A.1.1 Selection of methods

When selecting methods, variation of *inter alia* the following aspects should be considered:

- sample preparation, for example grinding/milling, extraction techniques, solvents; digestion media/volumes/conditions, clean-up steps;
- sample introduction and/or separation, for example using LC or GC;
- quantification principles, for example molecular or atomic absorption, mass spectrometry, flame ionization, fluorescence, etc.;
- calibration procedures, unless one approach has clear advantages, because of its metrological rigour or because of achieving lower measurement uncertainties.

In many cases, variation of all aspects will be impossible. In these cases, the maximum possible variation shall be sought. For example, if gas chromatography is the only available separation technique, then the study should at least aim to include different injection techniques, different columns and temperature programs and quantification by different detectors.

The producer should require that all methods used in the campaign are properly validated and that a reasonable estimate of the measurement uncertainty can be provided.

To allow laboratories the free choice of methods while ensuring the necessary variation, the producer should obtain information about the methods applied by the participants before the start of the study to obtain the necessary range of methods. If required, a targeted search for laboratories offering specific methods should be performed to avoid receiving results obtained by one method.

##### A.1.2 Selection of methods

An important decision is whether all laboratories should use the same calibrator or whether laboratories should be given free choice of the calibrator. Using a single calibrator reduces variation caused by different calibrators from different suppliers. On the other hand, any bias in this single calibrator will translate into the same bias in the certified values. Therefore, use of a single calibrator requires very careful characterization of this calibrator, requiring in many cases independent confirmation of purity or composition. Allowing laboratories the free choice of calibrator is logistically easier, does not require tests for independent confirmation and allows laboratories to apply their procedures unchanged, but means that some way of checking the appropriateness of the calibration should be included in the study.

As a general guideline:

- for well-established measurements, where experience shows that the quality of available calibration standards is sufficient, giving laboratories the free choice of standards is usually preferable;
- in cases where there is significant doubt about the quality of standards on the market, the efforts to characterize a common standard are often justified.

In some cases, different producers of commercial standards obtain their pure material from the same company. It is therefore useful to establish whether the original sources differed.

##### A.1.3 Selection of laboratories

Laboratories shall be selected based on demonstrated competence. Appropriate evidence for the demonstration of competence may include the following:

- results from previous proficiency tests;
- results on independent CRMs (possibly distributed as quality control materials together with the candidate CRM);
- method validation data;
- a full and credible uncertainty budget;
- previous participation in other RM certification campaigns for the same measurand; and

- third party assessment of compliance with ISO/IEC 17025 <sup>[53]</sup> or other relevant standards for the determination of the measurand in question.

The performance of externally assessed laboratories can differ in the same way as other competent laboratories. It is therefore prudent to obtain information on performance in addition to evidence of third party assessment of compliance with ISO/IEC 17025.

It shall be demonstrated that the measurements in each laboratory were performed in accordance with ISO/IEC 17025. In particular, the provisions of ISO/IEC 17025 regarding competence of staff, calibration of equipment and authorization (official release for use in line with ISO/IEC 17025: 2005, 4.3.2.1 <sup>[53]</sup>) of methods and results should be met.

In the absence of independent assessment, information on the extent of the laboratory's quality systems should be obtained.

NOTE 1 Potential gaps in fulfilling the requirements of the respective standard can sometimes be filled by the RM producer (e.g. an RM producer may archive the laboratories' raw data, if they do not have an archiving system).

NOTE 2 Obtaining calibration certificates for each and every instrument is in many cases impractical. Agreement of results on quality control samples can be used as demonstration of sufficient calibration of all relevant input factors.

#### A.1.4 Number of independent data sets

The number of participating laboratories is less important than the number of independent data sets. A single laboratory may be able to provide several data sets, all obtained by independent methods. The remainder of the discussion focuses on data sets, regardless whether each data set was provided by a different laboratory, some laboratories provide more than one data set or all data sets are provided by the same laboratory.

Complete independence of results is difficult to achieve if measurements are performed in a single laboratory. The producer should critically review the variation of all critical steps as outlined in A.1.1 to check whether sufficient method variability is present to demonstrate absence of bias for each individual step. This includes checking whether the same critical chemicals, equipment, calibrators, etc. were used by the various methods. The organizer should also remind the laboratories at the onset of the study not to censor data, i.e. not suppress/change data from one method without notification after a crosscheck of results between the various methods.

The producer should set a documented minimum number of technically valid results for which value assignment will be considered. The number of data sets should be large enough to provide an adequately small uncertainty in the estimated value after allowing for the possibility of failure to report, exclusion of results for technical reasons and the statistical evaluation intended.

For interlaboratory studies, it is advisable to plan to have five or more independent sets of technically valid results on which to base a certified value.

The following considerations may change the number of data sets required in order to achieve the desired uncertainty for certified values.

- Uncertainty required: The uncertainty of assigned values usually decreases with the number of technically valid data sets, requiring more data sets for smaller uncertainties.
- Technical difficulty: The less well established or the more technically challenging a measurement is, the larger the between-data set variation can be expected to be.
- Likelihood of technically invalid results: Even experienced laboratories can deliver technically invalid results which cannot be used for certification. Such results are more likely for unfamiliar materials, new or modified measurement procedures, challenging measurements or unusual reporting requirements. The number of participants or (for single laboratory studies) independent measurements should be increased where the risk of technical errors is higher.
- Reproducibility/repeatability ratio: Where between-data set variation is known to be the primary source of variation, preference should be given to a larger number of data sets rather than higher number of replicates for each data set.
- Statistical evaluation: Different data treatment methods may require larger numbers of observations to provide good numerical stability and/or achieve sufficiently small uncertainty.
- Validation results from a CRM: Results of consistent high quality obtained from one or more CRMs, used for quality assurance at each laboratory, serve to demonstrate the validity of all data sets and to provide information on bias among laboratories and methods. This information allows application of Bayesian models of uncertainty estimation for which fewer data sets will suffice than in a strictly frequentist approach.

### A.1.5 Number of units and replicate determinations

The number of units of the candidate CRM sent to each participant, the number of replicate determinations performed by each participant and the condition of these determinations are determined by practical as well as evaluation considerations.

- In the case of breakage of a unit and, if results are averaged across units, partially compensates for between-unit heterogeneity. If the heterogeneity is large, single measurements on several different units are preferable to several replicate measurements on a single unit. If contamination, breakage or heterogeneity are not an issue, sending a single unit is sufficient.
- In the absence of reliable uncertainty estimations by the participants, requesting measurements under conditions of intermediate precision can provide an indication of the reliability of the participants' results or a check on the reliability of the participants' uncertainty estimates.

### A.1.6 text

Although the study only includes laboratories of demonstrated competence, inclusion of additional samples for quality control has been found to be highly beneficial. Results on these samples can identify technical problems and aid the technical evaluation.

- RMs, in particular natural matrix RMs and quality control (QC) materials may be used to demonstrate the validity of the measurement result when measured alongside the unknown to be characterized.
- Spiked materials, spiked blanks, etc., may be used to check parts of the measurement procedure or to assist in the process of assigning values to a material.
- Blank matrix materials, blank extracts, etc., may be used to demonstrate that the measurement method provides a result not significantly different from zero when the characteristic of interest is not present (as often done in composition measurements), or to establish a correction or correction factor together with the uncertainty of the correction factor.

NOTE Sometimes CRMs used for quality control in an interlaboratory study are supplied without the original label to avoid identification. However, as the number of reliable CRMs is limited, experienced laboratories often recognize the material from the visual appearance and/or the values.

### A.1.7 text

The use of preformatted reporting forms can be useful as it has the advantage of structuring the report and (if transmitted electronically) allows copying of the results, which can reduce transcription errors in the RM producer's collation of results. Disadvantages of preformatted reporting forms are that they often force laboratories to depart from their reporting practices, which can lead to transcription errors. If reports are submitted electronically, the requirements of ISO Guide 34 [2] on the integrity of electronic records, especially of reports of test results shall be adhered to.

NOTE When reports are submitted in spreadsheet form, unintentional alteration can often be prevented by the use of 'locking' or 'protection' facilities incorporated in the spreadsheet software.

Laboratories should be requested to report individual results (not only averages over all samples), regardless of whether an uncertainty statement is reported or not, although reporting of an average and an expanded uncertainty and its coverage factor may be sufficient.

Where there is an option for correction of a known procedural bias, such as extraction recovery, the producer should state clearly whether results should be corrected or not. The producer may also require participants to report bias checks (e.g. from spikes) and use these to correct results for detected bias. Where a correction is applied by the participant, any reported uncertainty should include the uncertainty associated with the correction.

Participants should be instructed on how to report results near detection limits, if such results are likely to occur. Results reported as "less than" make statistical evaluations more difficult. On the other hand, reporting of results near detection limits contradicts many laboratories' quality procedures. Where results near detection limits are likely, producers should either require laboratories to report the observed (including negative) results instead of, or in addition to, their normal reporting, or should adopt statistical procedures that allow for "left-censored" results such as 'less than' statements or results restricted to values above zero.

It is recommended that an outline of the measurement procedure used is reported in sufficient detail to permit an understanding of all stages in the measurement process (e.g. in chemical analysis, the digestion/extraction of the sample and separation of the analytes of interest, clean-up, quantification).

Participants should be requested to give literature references, where applicable.

### A.1.8 Instructions for participants

Guidelines to participants should contain:

- a) a clear outline of the goal of the study;
- b) instructions to refrain from comparing results with other participants, including the reasons for discouraging collusion (that is, cooperative exchange of information);
- c) the number of units to be tested;
- d) the number of replicate determinations to be performed;
- e) the measurement methods to be used;
- f) the test portion size;
- g) requirements with respect to quality and traceability of the measurement results;
- h) the time schedule (distribution of samples, delivery of results);
- i) the mode of dispatch;
- j) instructions for intermediate storage of samples;
- k) specific instructions for sample treatment, if applicable;
- l) description of the mechanisms for verifying assumptions about the data; and if applicable, a statement that results will only be disseminated in an anonymous manner (i.e. no link between a certain laboratory and the results) or that the measurement task is regarded as fulfilled, even if results of a certain method are not used for technical reasons.

A meeting with the laboratories/groups involved (prior to distributing the samples and performing the measurements) may help all parties involved to align all actions to be carried out during the collaborative study, and to discuss possible problems and/or pitfalls.

The organizer of a study should take into consideration that communication of results of different methods among laboratory staff is common and it can be useful to take steps to discourage collusion in addition to b) above within and between laboratories. For example, different labelling of materials for each participant can make it harder for participants to compare results.

## A.2 Evaluation

### A.2.1 Initial screening

The producer may contact participants to assist in the investigation of anomalies at any stage of the evaluation process. If a participant is contacted, it is recommended that initial contact should not specify the nature of the anomaly (for example, the direction of deviation of the results); rather, the participant should initially be invited to investigate and report any errors discovered.

If data sets from more than one measurement method are provided by a single laboratory, initial inspection should consider the data sets individually (that is, as if independently reported by different laboratories). The RM producer should nonetheless take account of all data sets submitted by a laboratory when drawing conclusions about which of the data sets to retain when anomalies are found.

### A.2.2 Technical evaluation

Further technical examination to identify potential problems may include (but is not limited to) grouping results by techniques (measurement method and principle, sample pre-treatment methods, etc.) or the calibration technique used. In addition, comparison of the expanded uncertainties with the confidence interval of the mean of the submitted results gives an indication whether the stated uncertainty is realistic, as the expanded uncertainty should be at least as large as the confidence interval.

A technically invalid result is not necessarily an outlier nor is every outlier necessarily technically invalid. A result may fall well within the range of valid results, even when it is evident that the conditions, under which the result was obtained, were not in good order. Technically invalid results shall be removed from the data set or corrected, by repeating the measurement if possible and/or necessary. Conversely, a result deviating significantly from all other results may be the only technically valid result of the data set.

### A.2.3 Statistical evaluation

#### A.2.3.1 General principles for statistical evaluation

Characterization by a collaborative study or by multiple measurement procedures aims at randomization of bias between data sets. Statistical evaluation typically assumes that the true value of a measurand corresponds to the true value of a population parameter, usually the population mean.

#### **A.2.3.2 text**

Finding appropriate estimators for the expected value is closely linked to the (either assumed or determined) underlying distribution of values.

Determining a probability density function from a data set requires significantly more data than can practically be obtained, whereas checking for consistency with an assumed distribution and calculating parameters from it (e.g. average of a normal distribution) is possible with the number of data sets usually available in characterization studies.

The producer shall therefore check whether there is evidence of deviation from the assumed distribution using, for example, visual methods (histograms, Kernel density plots, and normal probability (or, more generally, Q-Q plots) or statistical checks for departure from particular distributions, including tests for normality or determination of skewness and kurtosis. An approximately normal distribution of data sets is often observed for results well above the limit of quantification; other distributions include Poisson distributions (e.g. microbe counts) or a Weibull distribution (e.g. mechanical failure of ceramics). The selected distribution shall be in agreement with the reported data as well as with the theoretical and historical knowledge of the measurement in question. If these differ significantly, no value shall be assigned unless technical reasons for the unexpected distribution can be given.

**EXAMPLE** Theoretical considerations as well as a multitude of interlaboratory comparisons indicate that results for trace elements in soil follow normal distributions (unless the level is close to the limit of quantification). Deviation of the observed data from normal distribution (e.g. tailing) indicates technical problems. As it is not clear whether the problem lies with the majority or with the tails, no value is assigned.

In some cases, the results can be transformed so that they become approximately normally distributed. Some commonly used transformations include logarithmic, square root and exponential forms. There should be a technical basis for such a transformation, as deviation of the results from the expected distribution may indicate technical problems.

#### **A.2.3.3 Outliers**

Outlying results can occur at all levels of a collaborative study: single observations, subgroups of observations (e.g. grouped per bottle), or the results from complete methods/laboratories can be observed to be outlying. Appropriate outlier tests for means and variance as well as Mandel's H and k statistics <sup>[24]</sup> may be used to identify outlying results.

Outlying observations or mean values should not be removed solely on the grounds of a statistical outlier test, but may be removed if there is a technical reason to do so.

**NOTE 1** Technical reasons include inadequate calibration, inadequate measurement procedures, use of inadequate reagents, failure to account for interferences, deviation from the certified value of an independent quality control material, etc.

**NOTE 2** Outlier tests usually ignore measurement uncertainties. An outlying data set may agree with the other data when the respective uncertainties are taken into account.

**NOTE 3** Data points with unusually large uncertainty can be removed on technical grounds if they agree with the other results within their uncertainty and if the uncertainty is so large as to indicate a technical problem or can be retained, as there is agreement between results.

**NOTE 4** Data sets that show a high outlying variance can indicate a lack of method repeatability or lack of intermediate precision, which may justify rejection on technical grounds.

**NOTE 5** Different methods may differ in precision and it can be important to retain data sets for a relatively imprecise method in order to retain a representative collection of methods

A special case of outlying variance is a result set having zero variance. Often, laboratories will set restrictions on the number of decimal places reported. Such a restriction may conflict with the purposes of the project. It is advisable to anticipate this problem in the instructions for reporting. Alternatively, the lab may be able to provide additional digits on request.

#### **A.2.3.4 Robust statistics**

Robust statistics provide a large collection of statistical methods that explicitly allow for the presence of outlying values in an otherwise approximately normally distributed data set. Typically, robust methods assign weights that decrease with the distance from the main body of the data. Robust methods exist for estimating the mean and standard deviation for simple outlier-contaminated data sets, as well as for many other parameters of interest.

Robust methods also exist for other situations including, for example, sets of reported values with appreciably different uncertainties, or the analysis of data consisting of multiple replicates for each data set.

Robust estimators provide high resistance to the influence of extreme outlying values. For symmetric 'heavy-tailed' distributions (that include a larger proportion of values far from the mean than would be expected from a normal distribution), robust statistics typically provide unbiased estimates of the mean with lower variance than the simple arithmetic mean. This approach results in an unbiased estimate with smaller uncertainty than the arithmetic mean would give in the same circumstances. For data with extreme outliers, robust statistics can be a very considerable improvement over the mean. A short summary of useful robust statistical methods, together with conditions for their use in RM characterization and references to more detailed descriptions, is given in Clause B.5.

**NOTE** The tolerance of robust statistics to extreme values additionally makes it useful for identifying outliers in larger data sets containing several extreme values.

### A.2.3.5 Clustering

If the results form two or more distinct clusters, the following possibilities should be considered.

- a) If there is correlation of these clusters with measurement methods or procedures (or if the only deviating value comes from a different method principle), and if the difference between the means of these clusters is statistically significant and if the difference between the means of these clusters is too large to permit a sufficiently small uncertainty for the intended use of the material, then no single property value can be assigned. In this case, improvement in the measurement method procedures is necessary to resolve the problem; otherwise, several procedurally defined property values could be assigned for each procedure.
- b) If there is correlation of these clusters with measurement methods or procedures, but the difference between the means of these clusters is relatively small, one single value may be assigned. An additional uncertainty term accounting for the between-method variation should be added to the uncertainty of characterization. There are several approaches to this estimation problem described in the literature.<sup>[54] [55]</sup>
- c) If there is no correlation of these clusters with measurement method procedures, and if the difference between these clusters is large, no value can be assigned. A larger pool of results may be necessary to overcome the relatively poor agreement of measurement methods and procedures available.

## A.3 Use of collaborative studies for multiple purposes

**A.3.1** ISO 5725-3<sup>[14]</sup> lists a number of purposes of interlaboratory comparisons, including value assignment of reference materials, evaluation of the performance of laboratories and evaluation of the performance characteristics of a method. In general, a particular study is best used for only one of these purposes; to do otherwise may compromise one objective in favour of another, or confuse the purpose of the study for the participants.

**A.3.2** The main advantage in combining characterization studies for RMs with other studies such as method performance studies and proficiency testing (PT) is an apparent reduction of costs.

**A.3.3** Potential incompatibilities or conflicts when combining PT or method performance studies with characterization studies include the following.

- a) PT studies aim to assess the competence of a laboratory whereas participation in characterization studies is restricted to laboratories of demonstrated competence. A combination of a PT study with a characterization study might therefore assume the very fact of demonstrated competence for the purpose of characterization that should be assessed in the PT part.
- b) The requirements of proficiency testing do not normally require extensive replication and rarely permit specification of equal replication by all participants.
- c) PT results are typically treated as confidential and access to details of experimental methods cannot be guaranteed.
- d) Participants often pay for participation in a PT study and the motivation for participation can sometimes be commercially driven. Participants are therefore often unwilling to adhere to the detailed study setup, requirements for traceability and quality assurance, calibration, provision of uncertainties, etc. required in a characterization study.
- e) Participation in PT studies is a requirement for accreditation and results are used for demonstration of competence to potential customers. Participants therefore may be less willing to discuss freely and to admit to technical problems required in a characterization study for the technical evaluation of results.

- f) PT studies usually assess laboratory performance in comparison with other laboratories or with an externally set criterion, whereas a characterization study aims at getting a good estimate of the true value and its uncertainty. This leads to different approaches in evaluation, especially in dealing with extreme values/outliers.
- g) PT studies are typically open to any laboratory willing to participate. Producers therefore often have little control over the methods applied by the participants, which can result in receiving results from only one method even if other methods exist.
- h) While feedback on the results for the purpose of checking anomalies cannot commence before a PT study closes, this information is required for following up apparent anomalies of results in a characterization study
- i) Method performance studies evaluate the performance of a method immediately after development, so there is little prior information on typical performance.
- j) The evaluation of a PT study also focuses very much on statistics, whereas characterization studies place more emphasis on the technical evaluation.
- k) Method performance studies typically apply to only one measurement procedure and are therefore not suited for assigning values to method-independent analytes.

**A.3.4** Because of these potential conflicts, consensus values from a PT scheme should not be used to certify a reference material without critical evaluation of the data. However, data collected from participants in a PT scheme may be used to certify a reference material provided that all of the requirements given below are met:

- a) a decision before the start of a study which subset of data from specified laboratories would be used for value assignment;
- b) demonstration of competence of these laboratories independently of this study, e.g. from performance in previous rounds of the scheme or by other means;
- c) organization of the study for these selected laboratories according to the criteria described in A.1.3;
- d) evaluation of the results of these laboratories according to the criteria laid out in Clause A.2;
- e) notification to proposed participants that results may be used in a certification study and permission obtained for access to technical detail.
- f)

## Příloha B (normativní)

### Statistické přístupy

#### B.1 Jednofaktorová analýza rozptylu (ANOVA)

Uvažujme případ, kdy existuje  $a$  skupin a každá z nich obsahuje  $n_i$  členů. V ideálním případě, počet členů ve skupinách by měl být stejný, ale v praxi tomu vždy tak není; některá data mohou z nejrůznějších důvodů „scházet“. Jsou vyvinuty výrazy, které počítají s těmito chybějícími daty a jsou doporučeny pro zpracování nekompletních souborů dat. Samozřejmě, čím méně úplné jsou soubory dat, tím horší je kvalita odhadů (hodnot).

Jednofaktorová ANOVA aplikovaná na studie referenčních materiálů je založena na statistickém modelu

$$x_{ij} = \mu + \delta_i + \varepsilon_{ij} \quad (\text{B.1})$$

kde

- $x_{ij}$  je výsledek pozorování  $j$  ve skupině  $i$ ;
- $\mu$  je pravá střední hodnota (průměr) základního souboru možných výsledků (ze kterého pozorované výsledky  $x_{ij}$  pochází);
- $\delta_i$  je (pravá) odchylka průměru skupiny  $i$  od  $\mu$ ;
- $\varepsilon_{ij}$  je náhodná chyba pozorování  $j$  ve skupině  $i$ .

předpokládá se, že  $\delta_i$  vzniká z výskytu výsledků (obvykle normálně rozdělených) s průměrem nula a směrodatnou odchylkou  $\sigma_{between}$  a o  $\varepsilon_{ij}$  se předpokládá, že vzniká z výskytu výsledků (opět obvykle normálně rozdělených) s průměrem nula a směrodatnou odchylkou  $\sigma_{within}$ . Předpokládá se, že pravá směrodatná odchylka každé skupiny je stejná pro všechny skupiny. Tyto dvě směrodatné odchylky jsou odpovědné za distribuci pozorovaných dat. Úkolem zde aplikované ANOVA je zjistit velikost těchto dvou směrodatných odchylek.

Rozptýlení dat může být vyjádřeno formou součtů druhých mocnin odchylek (rozdílů), známé také jako „součty čtverců“. Tyto součty čtverců vyjadřují rozptýlení dat na různých hierarchických úrovních při analýze rozptylu [56]. Tak zvané hodnoty průměrných součtů čtverců odchylek uvnitř skupin resp. mezi skupinami,  $M_{within}$  a  $M_{between}$ , získané např. ze statistického programu softwaru počítače, mohou být převedeny (přepočítány) na odhadované rozptyly následovně:

$$s_{within}^2 = M_{within} \quad (\text{B.2})$$

$$s_{between}^2 = \frac{M_{between} - M_{within}}{n_0} \quad (\text{B.3})$$

kde

$$n_0 = \frac{1}{a-1} = \left[ \frac{\sum_i n_i - \frac{\sum_i n_i^2}{a}}{\sum_i n_i} \right] \quad (\text{B.4})$$

a  $s_{between}$  a  $s_{within}$  jsou odhady  $\sigma_{between}$  a  $\sigma_{within}$ , v uvedeném pořadí.

Jestliže ve studii v níž se plánuje, že každá skupina zahrnuje  $n$  pozorování, nescházejí žádná data, potom  $n_0$  je rovno  $n$ .

Tam, kde je to nezbytné, úhrnný průměr  $\bar{x}$  se může zjistit ze vztahu

$$\bar{x} = \frac{1}{\sum_{i=1}^a n_i} \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij}$$



(B.5)

Při absenci jakéhokoliv vlivu mezi skupinami,  $s_{between}$  je nula (nebo se blíží nule). Jestliže, z experimentálních důvodů, se pro  $s_{between}^2$  získá záporná hodnota, pak by se měla nastavit na nulu.

POZNÁMKA 1. Když je původní odhad  $s_{between}^2$  záporný,  $s_{between} = 0$  a  $s_{within} = s(x)$ , kde  $s(x)$  je směrodatná odchylka všech pozorování  $x_{ij}$ , toto vytváří maximálně pravděpodobné odhady těchto dvou parametrů (shodně s omezeným maximálně pravděpodobným odhadem). Představuje to též statisticky platný postup.

POZNÁMKA 2. Při studiu homogenity mezi jednotkami,  $s_A$  se rovná směrodatné odchylce  $s_{bb}$  mezi jednotkami a pozorování na každé jednotce RM tvoří 'skupinu'.

## B.2 Smíšené náhodné vlivy při analýze dat: Dvufaktorová ANOVA

Tento model se používá, když výsledky měření z mezilaboratorní spolupráce jsou použity současně k potvrzení homogenity materiálu a rovněž k jeho charakterizaci, a když je opakovatelnost ve skupinách stejná. Experimentální uspořádání je znázorněné na obr.B.1 pro případ mezilaboratorní studie. Jestliže se studie skládá z různých metod, její uspořádání je v podstatě stejné.

Výsledky mohou být popsány rovnicí modelu (B.6.):

$$x_{ijk} = \mu + A_i + B_{ij} + \varepsilon_{ijk} \quad (\text{B.6})$$

kde

- $x_{ijk}$  je  $k$ -tý výsledek jednotky vzorku  $j$  prezentovaný metodou/laboratoří  $i$ ;
- $\mu$  je pravá střední hodnota (průměr) základního souboru možných výsledků (ze kterého pozorované výsledky  $x_{ij}$  pochází);
- $A_i$  je chyba způsobená metodou/laboratoří  $i$ ;
- $B_{ij}$  je chyba způsobená  $j$ -tým vzorkem v rámci metody/laboratoře  $i$ ;
- $\varepsilon_{ijk}$  je náhodná chyba pro pozorování  $k$  ve skupině  $ij$ .

$A_i$  je odvozeno z výskytu výsledků s průměrem nula a směrodatnou odchylkou  $\sigma_L$ ,  $B_{ij}$  z výskytu výsledků s průměrem nula a směrodatnou odchylkou  $\sigma_{bb}$  a  $\varepsilon_{ijk}$  z výskytu výsledků s průměrem nula a směrodatnou odchylkou  $\sigma_r$ . Tak jako dříve se předpokládá, že člen vyjadřující chybu měření má stejný rozptyl nejenom uvnitř každé jednotky RM, ale také uvnitř každé laboratoře. Obvykle se také předpokládá, že výskytu výsledků jsou normálně rozděleny. Získanými parametry jsou odpovídající odhady směrodatné odchylky mezi laboratořemi  $s_L$ , směrodatná odchylka mezi jednotkami  $s_{bb}$ , a směrodatná odchylka opakovatelnosti  $s_r$ . Tyto jsou vztaheny k výrazům vyjadřujícím chybu následujícím způsobem:

$$s_L = \sqrt{\text{Var}(A_i)}$$

$$s_{bb} = \sqrt{\text{Var}(B_{ij})}$$

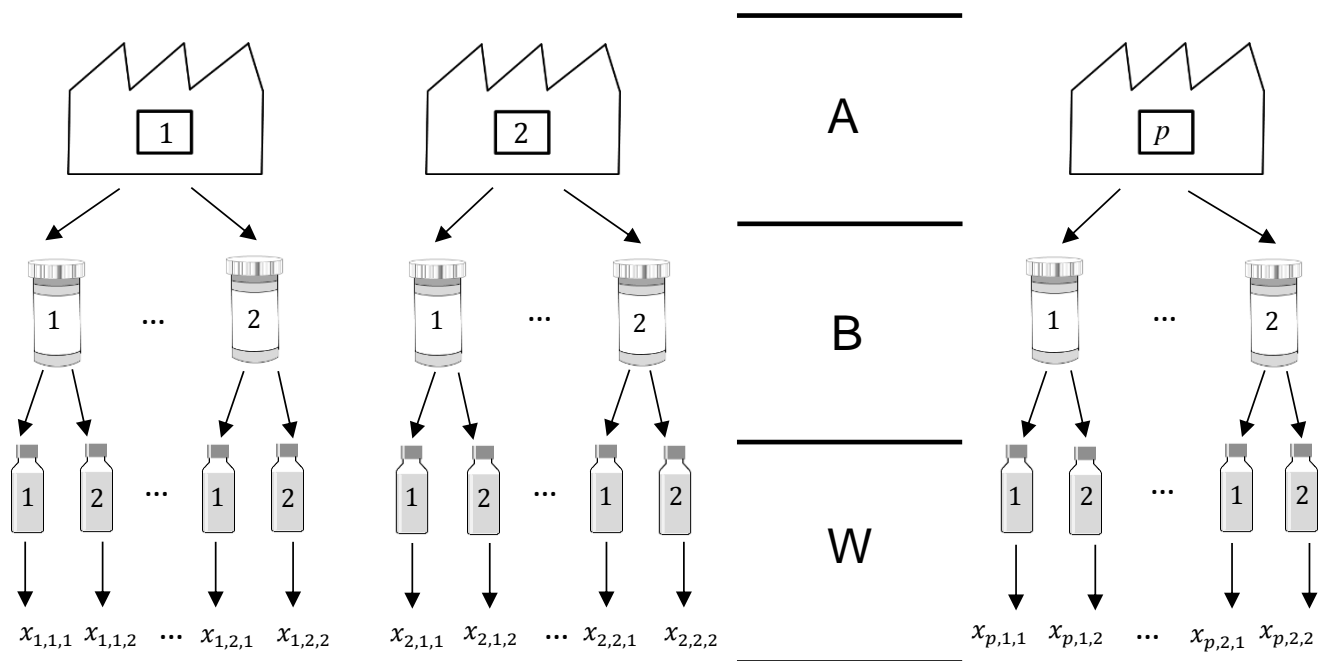
$$s_r = \sqrt{\text{Var}(\varepsilon_{ijk})} \quad (\text{B.7})$$

Pro homogenitu mezi jednotkami ( $s_{bb}$ ) platí stejné úvahy vzhledem k nemožnosti určit heterogenitu šarže, stejně jako je tomu při samostatné studii homogenity (viz 7.8).

Všechny tyto parametry mohou být zjištěny současně analýzou rozptylu (ANOVA), [56], jestliže je k dispozici dostatečný počet stejně replicitních výsledků (stejný počet opakovaných stanovení z každé jednotky vzorku a stejný počet jednotek pro každou metodu/laboratoř) po vyloučení všech technicky nebo statisticky neplatných výsledků. Jestliže požadavky ANOVA nemohou být splněny v důsledku značného množství neplatných a/nebo scházejících výsledků, platnost rozptylu mezi jednotkami by měla být zjištěna jiným způsobem (viz odstavec 7).

Teoretické detaily a další metody pro vyváženou a nevyváženou ANOVA jsou uvedeny v zavedených publikacích (viz např. [(58), [59][60], [61]).

Diskuse o analýze rozptylu (ANOVA) v souvislosti s certifikací referenčních materiálů jsou uvedeny v následující literatuře (viz např. [16], [56], [62], [63]).

**Klíč**

- A variabilita mezi laboratořemi  
 B variabilita mezi jednotkami (velké nádoby vyznačují jednotky RM)  
 W opakovatelnost měření (malé nádoby označují jednotlivé podíly vzorku)

**Obrázek B.1 — Náčrt mezilaboratorní studie spojené se studií homogenity šarže [charakterizace RM (dvoufaktorové uspořádání)]**

Výrazy pro výpočet  $s_L$ ,  $s_{bb}$  a  $s_r$  jsou uvedeny dále (viz Reference [16],[13],[64]). Úhrnný průměr  $\bar{x}$  se vypočítá podle vztahu

$$\bar{x} = \frac{1}{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij}} \sum_{i=1}^p \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} x_{ijk} \quad (\text{B.8})$$

kde  $p$  značí počet laboratoří,  $b_i$  je počet jednotek vzorku použitý metodou/laboratoří  $i$  a  $n_{ij}$  je počet opakovaných měření jednotky vzorku  $ij$ .  $s_L$ ,  $s_{bb}$  a  $s_r$  jsou vyjádřeny pomocí vztahů

$$s_r^2 = M_{\text{within}} \quad (\text{B.9})$$

$$s_{bb}^2 = \frac{M_B - M_{\text{within}}}{n_0} \quad (\text{B.10})$$

$$s_L = \frac{M_L - n'_0 s_{bb}^2 - M_{\text{within}}}{n_0} \quad (\text{B.11})$$

Průměry čtverců ( $M$ ) lze získat pomocí publikovaných metod (např. reference [16] a [64] nebo pomocí vhodného statistického softwaru. Uvedené výrazy platí pro chybějící nebo vyřazená (neplatná) data. Pro kompletní soubory dat je možné použít jednodušší výrazy uvedené v ISO 5725-3 [14].

Výpočet  $s_{bb}$  uvedený zde je možné použít pro odhad nejistoty spojený s heterogenitou stejným způsobem tak jak je vypočteno v kapitole 7.

Pokud je použit v souvislosti s mezilaboratorní studií, je velmi důležité ověřit platnost předpokladu o stejném rozptylu mezi jednotlivými skupinami. Tento předpoklad může být ověřen pomocí statistických testů, jako jsou Bartlettův test a Levenův test [58],[64].

### B.3 Lineární regrese (univariátní lineární model)

#### B.3.1 Základní model

Ve studiích o referenčních materiálech se lineární regrese používá ke stanovení jednoduchých poměrů změn a testování jejich statistické významnosti. Tato kapitola popisuje nejjednodušší případ. Jednoduchý lineární model [57], může být vyjádřen vztahem

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (\text{B.12})$$

kde  $\beta_0$  a  $\beta_1$  jsou úsek a směrnice (úsek kalibrační přímky a směrnice kalibrační závislosti) a  $\varepsilon$  značí složku náhodné chyby o níž se obvykle předpokládá, že je normálně rozdělena s průměrem nula.

Mějme sérii dvojic  $n$  pozorování  $Y$  versus  $X$ ; vzájemný vztah jednotlivých pozorování  $(x_i, y_i)$  je vyjádřen vztahem

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (\text{B.13})$$

**PŘÍKLAD** V základní studii stability,  $x_i$  odpovídá pozorovanému času a  $y_i$  odpovídá pozorované hodnotě sledované vlastnosti.

#### B.3.2 Fitování modelu

Regresní parametry mohou být vypočítány z následujících výrazů:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{B.14})$$

$$b_0 = \bar{y} - b_1 \bar{x} \quad (\text{B.15})$$

kde  $b_0$  a  $b_1$  jsou odhadované úsek a směrnice a  $\bar{x}$  a  $\bar{y}$  jsou průměry příslušných pozorování. Směrodatné chyby  $s(b_1)$  a  $s(b_0)$  se mohou vypočítat z následujících vztahů:

$$s(b_1) = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (\text{B.16})$$

kde

$$s^2 = \frac{\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2}{n - 2} \quad (\text{B.17})$$

a

$$s(b_0) = s(b_1) \sqrt{\frac{\sum_{i=1}^n x^2}{n}} \quad (\text{B.18})$$

### B.3.3 Kontrola a ověření předpokladů

Typické přezkoumání předběžně upraveného modelu zahrnuje:

- výpočet reziduí [ $r_i = (b_0 + b_1 x_i)$ ] a jejich vynesení proti času, následované kontrolou přítomnosti zakřivení, vlivy mezi jednotlivými sériemi měření, podstatné rozdíly v rozptýlení od jednoho časového bodu k druhému, nebo odlehle hodnoty;
- příprava grafů normality reziduí (nebo „Q-Q“ grafů) a kontrola přítomnosti ne-normálního rozdělení;
- alternativně, testy významných rozdílů v rozptýlení, významných rozdílů mezi skupinami reziduí a nelinearity.

### B.3.4 Testování statisticky významné změny

V tomto jednoduchém případě, po potvrzení platnosti základních regresních předpokladů, je obvyklým testem pro statisticky významný gradient  $t$ -test pro směrnici, která se významně liší od nuly. Ten se provádí výpočtem  $t$  statistiky

$$t_{b_1} = \frac{|b_1|}{s(b_1)} \quad (\text{B.19})$$

a jeho porovnáním s dvoustrannou kritickou hodnotou Studentova  $t$  pro  $n-2$  stupňů volnosti na 95% hladině spolehlivosti. Jestliže vypočítané  $t_{b_1}$  převyšuje kritickou hodnotu, považuje se směrnice za významně se lišící od nuly na 95% hladině spolehlivosti.

### B.3.5 Konfidenční interval pro regresní přímku

Nalezené hodnoty  $b_0$  a  $b_1$  se používají k odhadu  $\hat{y}$ , hodnoty odezvy pro příslušnou hodnotu  $\hat{x}$  podle vztahu

$$\hat{y} = b_0 + b_1 \hat{x} \quad (\text{B.20})$$

Často je užitečné získat konfidenční interval pro predikovanou hodnotu  $\hat{y}$ . Dvoustranný konfidenční interval pro 95% hladinu spolehlivosti je vyjádřen vztahem

$$\hat{y} \pm t_{95, n-2} s \sqrt{\frac{1}{n} + \frac{(\hat{x} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (\text{B.21})$$

kde  $t_{95, n-2}$  je dvoustranná hodnota Studentova  $t$  pro  $n-2$  stupňů volnosti. Konfidenční intervaly spolehlivosti pro různé hladiny spolehlivosti se podobně získají použitím příslušných kritických hodnot Studentova  $t$ .

POZNÁMKA Tento interval není predikčním intervalem pro jednotlivé hodnoty  $y_i$ .

## B.4 Předpověď životnosti RM nebo nalezení počátečního monitorovaného bodu založené na výsledcích studie stability

### B.4.1 Princip

Tam, kde je akceptovatelný rozsah, popsáný horní nebo dolní mezí specifikace, je užitečné – pro sledované změny hodnoty vlastnosti s časem při skladování – zjistit buď ‚skladovací životnost‘ RM nebo nastavit počáteční monitorovací bod, užívající výsledků ze studie stability.

V principu se jedná o zjištění 95%ního konfidenčního intervalu pro budoucí hodnoty RM, berouce v úvahu rychlost degradace a nalezení nejkratšího času, při kterém jedna z konfidenčních mezí (horní nebo spodní) se protne s mezí specifikace. Tento princip je schematicky ilustrován na obr. B.2.

Pro CRM, jakýkoliv nalezený a akceptovatelný rozsah pro certifikovanou hodnotu musí být menší než rozšířená nejistota, ideálně menší než  $U_{CRM}/3$ . Pro necertifikované referenční materiály jakákoliv akceptovatelná změna musí být dostatečně malá, aby se předešlo nežádoucím vlivům na zamýšlené použití.

### B.4.2 Předpověď životnosti v případě lineárního trendu

Mějme horní a dolní limit akceptovatelného rozsahu hodnot vlastností RM,  $L_{upr}$  resp.  $L_{lwr}$ . Konfidenční interval okolo předpokládané budoucí hodnoty RM – pokud degradace probíhá lineárním způsobem - je dán výrazy (B.20) a (B.21), kde  $x$  je čas od počátku studie stability. Požadovaná životnost se získá použitím výrazu (B.21) pro každou mez specifikace a řešením pro  $\hat{x}$  tak, aby se získala nejkratší kladná životnost  $t_{shelf}$ . To znamená, že  $t_{shelf}$  se zvolí podle výrazu (B.22)

$$t_{shelf} = \min(t_{s,upr}, t_{s,lwr}) \quad (B.22)$$

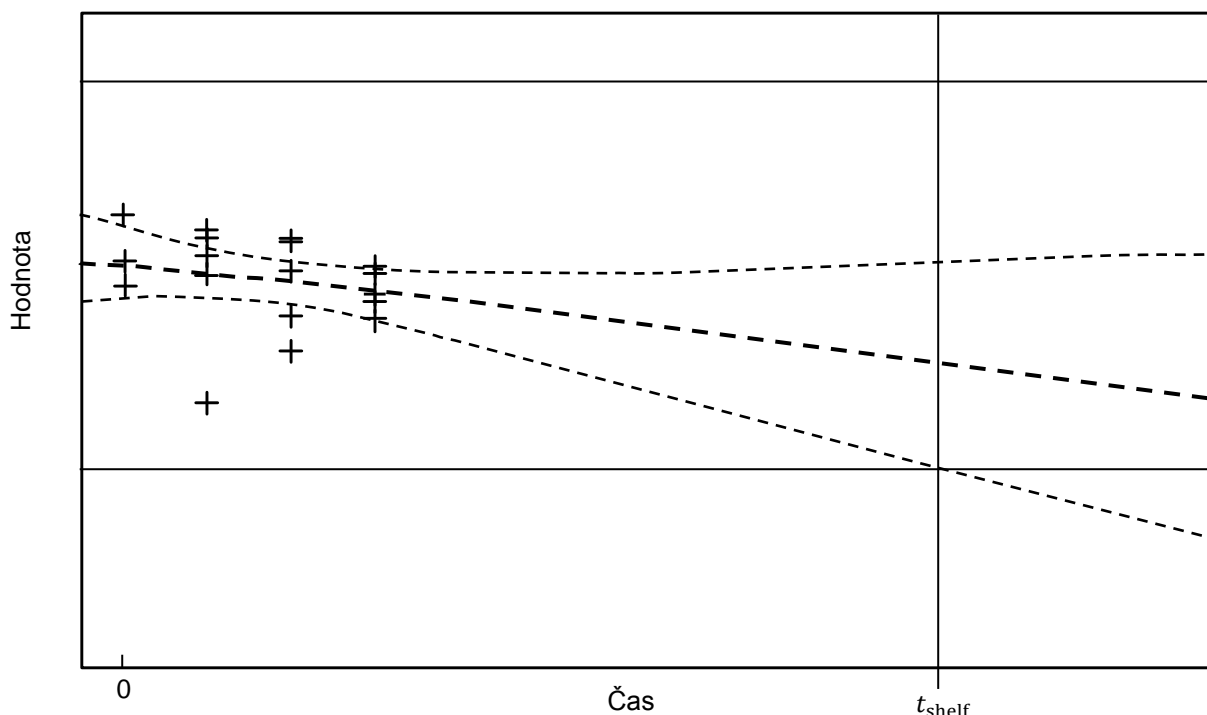
kde  $t_{s,upr}$  a  $t_{s,lwr}$  jsou nejnižší kladná řešení pro

$$L_{upr} - \left[ b_0 + b_1 t_{s,upr} \pm t_{95,n-2^S} \sqrt{\frac{1}{n} + \frac{(t_{s,upr} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right] = 0 \quad (B.23)$$

$$L_{lwr} - \left[ b_0 + b_1 t_{s,lwr} \pm t_{95,n-2^S} \sqrt{\frac{1}{n} + \frac{(t_{s,lwr} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right] = 0 \quad (B.24)$$

POZNÁMKA 1. Výrazy (B.23) a (B.24) mají celkem čtyři reálná řešení. Tam, kde neexistuje kladné řešení pro jednu z těchto rovnic – např. pro statisticky významnou degradaci, při které použitím výrazu (B.24) vznikají dvě kladná řešení – negativní řešení nejsou zohledněna.

POZNÁMKA 2. Výrazy (B.23) a (B.24) nemohou být řešeny algebraicky; oba musí být řešeny iterativně, např. použitím „vyhledávací“ metody.



POZNÁMKA Křížky znázorňují údaje ze studie stability, začínající v čase nula. Plné vodorovné čáry ukazují horní a dolní specifikační limit (mez). Rovná čárkovaná čára je čára získaná metodou nejmenších čtverců postupem popsáným v kapitole B.3. Zakřivené čárkované čáry vyjadřují horní a dolní interval spolehlivosti pro čáru vypočtenou pomocí vzorce (B.21) Plná svislá čára znázorňuje hodnotu  $t_{shelf}$ .

### Obrázek B.2 — Princip výpočtu životnosti ze studie stability

POZNÁMKA 3 Jestliže jsou specifikační meze vzájemně příliš blízko (přesněji řečeno, jestliže každá specifikační mez leží uvnitř intervalu daného vzorcem B.21), není možné nalézt řešení s reálnou hodnotou.

#### B.4.3 Předpověď životnosti pro ostatní degradační modely

Pro různé degradační modely, životnost může být předpovězena na základě stejných obecných pravidel jak byly popsány v odstavci B.4.2., tj. s výpočtem konfidenčních intervalů a tudíž pomocí vztahů (B.23) a (B.24), nastavených tak, aby vyhovovaly příslušnému degradačnímu modelu.

#### B.4.4 Volba počátečního monitorovacího bodu

Stejný princip se může použít k nalezení počátečního bodu monitorování i při post-certifikačním sledování stability; nalezená akceptovatelná změna hodnoty vlastnosti a hodnota  $t_{shelf}$ , vypočtená jak bylo shora popsáno, potom ukazují čas od zahájení studie stability, od něhož by se monitorování mělo poprvé uskutečnit.

## B.5 Robustní statistika

### B.5.1 Robustní odhady v mezilaboratorních studiích

Řada základních robustních odhadů robustnosti vhodných pro mezilaboratorní studie je uvedena v tabulce B.1, spolu s literárními odkazy a některými klíčovými ukazateli jejich fungování. Jejich použití je podmíněno následujícím:

- každý laboratorní výsledek je reprezentován jednou hodnotou (která může být sama o sobě průměrem, mediánem nebo jiným místním odhadem);
- informace a výsledky z laboratoří jsou podrobně zkoumány na přítomnost odborných (technických) chyb a nalezené technicky neplatné výsledky jsou vyjmuty ze souboru dat ještě před použitím robustních metod;
- po odstranění všech výsledků z technických důvodů by nemělo zůstat méně než 10 dat (údajů);

- d) většina hodnot v souborech dat (obvykle jejich střední část) je přibližně normálně rozdělena – nebo alespoň symetricky rozdělena (po event. transformaci, je-li vyžadována);
- e) neexistuje významné vytváření skupin (výsledků) podle jednotlivých zkušebních metod, zeměpisného původu, atd.;
- f) podíl extrémně odlehých hodnot v souboru dat nepřevyšuje polovinu bodu selhání použitého odhadu

Robustní odhady uvedené v tabulce B.1. netvoří exkluzivní seznam. Algoritmus A a Hampelův odhad robustnosti mohou být vyladěny buď změnou ladících parametrů, nebo velkým množstvím existujících alternativních algoritmů. Kolísání uvedených algoritmů a jiné robustní odhady se používají jako prostředek k ukázkám podobné nebo lepší výkonnosti než je výkonnost u odhadů, uvedených v tabulce B.1. Výkonnost dalších odhadů se posuzuje z teoretického bodu selhání, relativní účinnosti a dalšími vlastnostmi (jako jsou odolnost vůči asymetrii, nebo výkonnosti při malých souborech dat) které mohou být vhodné pro danou aplikaci.

Počet údajů ad c), smí být pod hodnotou 10, jestliže výrobce RM může dostatečným způsobem prokázat výkonnost vybraného odhadu na souboru dat typickém pro danou aplikaci.

POZNÁMKA 1 Horní hranice pro podíl extrémních hodnot je záměrně položena značně pod bod selhání odhadu buď aby se předešlo nepříznivým vlivům z vyššího podílu odlehých hodnot a protože obvykle při překročení tohoto podílu je vhodné zvažovat, zdali by neměl být certifikační proces zrušen v důsledku nedostatku shody mezi laboratoři.

POZNÁMKA 2. Odhady uvedené v Tabulce B.1 nepoužívají prezentované nejistoty nebo informace o vnitrolaboratorní preciznosti. Tyto informace a jednoduchý aritmetický průměr jsou proto nevhodnější v případě, že nejistoty přidružené k průměrným laboratorním hodnotám nejsou buď k dispozici, nebo nejsou spolehlivé.

POZNÁMKA 3. Odlehlé výsledky mohou nepříznivě ovlivňovat některé statistické postupy (včetně analýzy rozptylu), které se používají k testování významných vlivů při vytváření skupin. Pro tento účel se jeví jako vhodné opakovat statistický test po vyjmutí extrémně odlehých výsledků, aby bylo možné posoudit vliv odlehých výsledků na použitý statistický test.

POZNÁMKA 4. Nejrobustnější odhady mohou být vyladěny ke zvýšené odolnosti vůči odlehlým výsledkům na úkor snížené relativní účinnosti změnou ladících parametrů. Nejlepší rovnováha mezi účinností (která ovlivňuje nejistotu) a odolností vůči odlehlým výsledkům závisí na distribuci dat, včetně dat odlehých. Typické jsou ladící parametry, které poskytují relativní účinnost mezi 0,80 a 0,95 pro normální rozdělení.

### ...Tabulka B.1 – Některé robustní odhady a jejich vlastnosti

Robustní metoda	Odhadovaný parametr základního souboru (Poznámka 1)	Relativní účinnost $e_R$ (Poznámka 2, Poznámka 3)	Bod selhání (Poznámka 4, Poznámka 5)	Odkaz (Poznámka 6)
střední hodnota	průměr	1,0	0 %	
medián	průměr	0,67	50 %	
algoritmus A	průměr a směrodatná odchylka	0,97 <sup>a,b</sup> , 0,73 <sup>c</sup>	25 %	ISO 5725-3, ISO 13528
Hampel (s $Q_n$ nebo MADe stupnicí)	průměr	0,96 <sup>d</sup>	50 %	ISO 13528
výběrová směrodatná odchylka	směrodatná odchylka	1,0	0 %	
normalizované mezikvartilové rozpětí (nlQR)	směrodatná odchylka	0,37	25 %	ISO 13528
škálová absolutní mediánová odchylka (MADe)	směrodatná odchylka	0,37	50 %	ISO 13528
$Q_n$	směrodatná odchylka	0,81	50 %	ISO 13528
POZNÁMKA 1 Přesně řečeno, zde zmíněný výskyt výsledků podléhá normálnímu rozdělení, tj. předpokládá se, že je kontaminován odlehlými výsledky.				

POZNÁMKA 2. Relativní účinnost je mírou rozkolísanosti odhadu, pokud je aplikován na normálně rozdělená data, tzn. rozptyl odhadu dělený aritmetickým průměrem nebo (pro odhady pomocí směrodatné odchylky) výběrovou směrodatnou odchylkou pro stejná data. Vysoká účinnost je lepší. Uvedené účinnosti jsou vysoké; pro uvedené odhady se relativní účinnost (ve srovnání s průměrem nebo směrodatnou odchylkou) mírně zlepšuje pro menší soubory dat.

POZNÁMKA 3. Zatímco robustní odhady nejsou tolik účinné pro normálně rozdělená data jako je průměr (relativní účinnost < 1,0), pro další rozdělení, s masivním „chvostováním“ nebo odlehlými výsledky, robustní odhady se rychle zlepšují (vůči aritmetickému průměru). To samé platí pro směrodatnou odchylku vzorku a odpovídající robustní odhad směrodatné odchylky.

POZNÁMKA 4. Bod selhání je mírou odolnosti odhadu k podílu extrémně odlehlých výsledků. Vyšší hodnota je lepší.

POZNÁMKA 5. Zde použitá definice bodu selhání je podíl velmi normálně rozložených souboru dat, který může být posunut do plus nekonečna aniž by se odhad také posunul do nekonečna. Například jestliže těsně méně než 50 % souborů dat je přesunuto do plus nekonečna, medián zůstane uvnitř zbývajících konečných dat.

POZNÁMKA 6. Jsou uvedeny odkazy na jiné ISO normy než pro průměr, medián a výběrovou směrodatnou odchylku, které jsou velmi dobře známy.

POZNÁMKA 7. Informace o výkonnosti uvedené v této tabulce pocházejí z normy ISO 13528:2015 [10].

- |   |  |
|---|--|
| a | pro parametr ladění, $k = 1,5$   |
| b | účinnost pro odhad algoritmu A průměru s parametrem ladění $k = 1,5$ jako v ISO 5725-3.                  |
| c | účinnost pro odhad algoritmu A směrodatné odchylky s parametrem ladění $k = 1,5$                         |
| d | účinnost pro Hampelův algoritmus s parametrem ladění $a = 1,5$ , $b = 3,0$ , $c = 4,0$ jako v ISO 13528. |

### B.5.2 Odhad nejistot robustních odhadů

Nejistota jakéhokoliv odhadu založeného na umístění odhadu odvozeného přímo z mezilaboratorních dat závisí částečně na použitém odhadu a částečně na distribuci, ze které jsou data sestavena. Pro normálně rozdělená data, statistická nejistota spojená s robustním odhadem je

$$\frac{S^*}{\sqrt{e_R N}} \quad (\text{B.25})$$

Robustní odhady jsou však zřídka kdy používány pro normálně rozdělená data; jsou používány pro data, která zahrnují vysoký podíl hodnot velmi vzdálených od (pravého) průměru a které mohou obsahovat mnoho extrémně odlehlých výsledků. Za těchto okolností, nejistota je vždy větší než nejistota poskytnutá ze vzorce (B.25). Nejvhodnější nejistota potom značně závisí na rozdělení dat. Hlavní metodou pro přidělení nejistot odhadům ze značně ne-normálních dat je simulace nebo opětovné vzorkování. Při absenci nejistot založených na simulačních studiích nebo převzorkování z určitého souboru dat, nejistota charakterizace pro robustní odhady by neměla být menší než

$$1,253 \frac{S^*}{\sqrt{N}} \quad (\text{B.26})$$

Koeficient 1,253 je odvozen ze směrodatné odchylky mediánu, účinnosti mediánu jako odhadu střední hodnoty pro rozsáhlý soubor výsledků odvozených z normálního rozdělení. Zatímco účinnost více sofistikovaných robustních metod může být větší než účinnost mediánu, tento koeficient se doporučuje jako konzervativní (vysoký) odhad, dovolující mírnou ne-normalitu.



## Annex B (normative)

### Statistical approaches

#### B.1 One-way analysis of variance (ANOVA)

Consider the case that there are  $a$  groups, and each of them contains  $n_i$  members. Ideally, the number of members in the groups should be equal, but in practice, this is not always the case; some data may be “missing” for a variety of reasons. Expressions have been developed to account for these missing data [56], [57] and are recommended for treating incomplete data sets; however, the more incomplete the data set becomes, the poorer the quality of the estimates.

One-way ANOVA applied to reference material studies is usually based on the statistical model

$$x_{ij} = \mu + \delta_i + \varepsilon_{ij} \quad (\text{B.1})$$

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  $i$ .

$\delta_i$  is assumed to arise from a population (usually normally distributed) with mean zero and standard deviation  $\sigma_{\text{between}}$  and  $\varepsilon_{ij}$  is assumed to arise from a population (again usually normally distributed) with mean zero and standard deviation  $\sigma_{\text{within}}$ . Note that the true standard deviation for each group is assumed to be the same for all groups. These two standard deviations account for the spread of data observed; the purpose of ANOVA as applied here is to estimate these two standard deviations.

The scattering of data can be expressed in terms of sums of squared differences, also known as “sums of squares”. These sums of squares express the scattering at various hierarchical levels in the analysis of variance.[56] The so-called within- and between-group mean squares  $M_{\text{within}}$  and  $M_{\text{between}}$ , respectively, as obtained from a spreadsheet or statistical software program, can be converted into the estimated variances as follows:

$$s_{\text{within}}^2 = M_{\text{within}} \quad (\text{B.2})$$

$$s_{\text{between}}^2 = \frac{M_{\text{between}} - M_{\text{within}}}{n_0} \quad (\text{B.3})$$

where

$$n_0 = \frac{1}{a-1} = \left[ \frac{\sum_i n_i - \frac{\sum_i n_i^2}{a}}{\sum_i n_i} \right] \quad (\text{B.4})$$

and  $s_{\text{between}}$  and  $s_{\text{within}}$  are the estimates of  $\sigma_{\text{between}}$  and  $\sigma_{\text{within}}$ , respectively.

When there are no missing data in a study planned to contain  $n$  observations per group,  $n_0$  becomes equal to  $n$ .

Where necessary, the grand mean  $\bar{x}$  may also be estimated from

$$\bar{x} = \frac{1}{\sum_{i=1}^a n_i} \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij}$$

(B.5)

In the absence of any between-group effect,  $s_{\text{between}}$  is expected to be (close to) zero. If, for experimental reasons, a negative value is obtained for  $s_{\text{between}}^2$ , then it should be set to zero.

NOTE 1 When the initial estimate of  $s_{\text{between}}^2$  is negative,  $s_{\text{between}} = 0$  and  $s_{\text{within}} = s(x)$ , where  $s(x)$  is the standard deviation of all the observations  $x_{ij}$ , from the maximum likelihood estimates of the two parameters (identical to the restricted maximum likelihood estimate). This is also a statistically valid procedure.

NOTE 2 In a between-unit homogeneity study,  $s_A$  is identical with the between-unit standard deviation  $s_{\text{bb}}$  and the observations on each RM unit form a 'group'.

## B.2 Nested random effects in data analysis: Two-way ANOVA

This model may be used when the results of the measurement campaign collaborative study are used to confirm the homogeneity of the material, as well as to characterize it, and where the repeatability is constant across groups. The experimental scheme is illustrated in Figure B.1 for the particular case of an inter-laboratory study. When a campaign consists of different methods, the layout for the campaign is essentially the same.

The results can be described by the model equation (B.6):

$$x_{ijk} = \mu + A_i + B_{ij} + \varepsilon_{ijk} \quad (\text{B.6})$$

where

- $x_{ijk}$  is the  $k$ 'th result of sample unit  $j$  reported from method/laboratory  $i$ ;
- $\mu$  is the true mean value of the population of possible results (from which the observed results  $x_{ijk}$  are assumed to arise);
- $A_i$  is the error due to method/laboratory  $i$ ;
- $B_{ij}$  is the error due to the  $j$ 'th sample unit within method/laboratory  $i$ ;
- $\varepsilon_{ijk}$  is the random error for observation  $k$  in group  $ij$ .

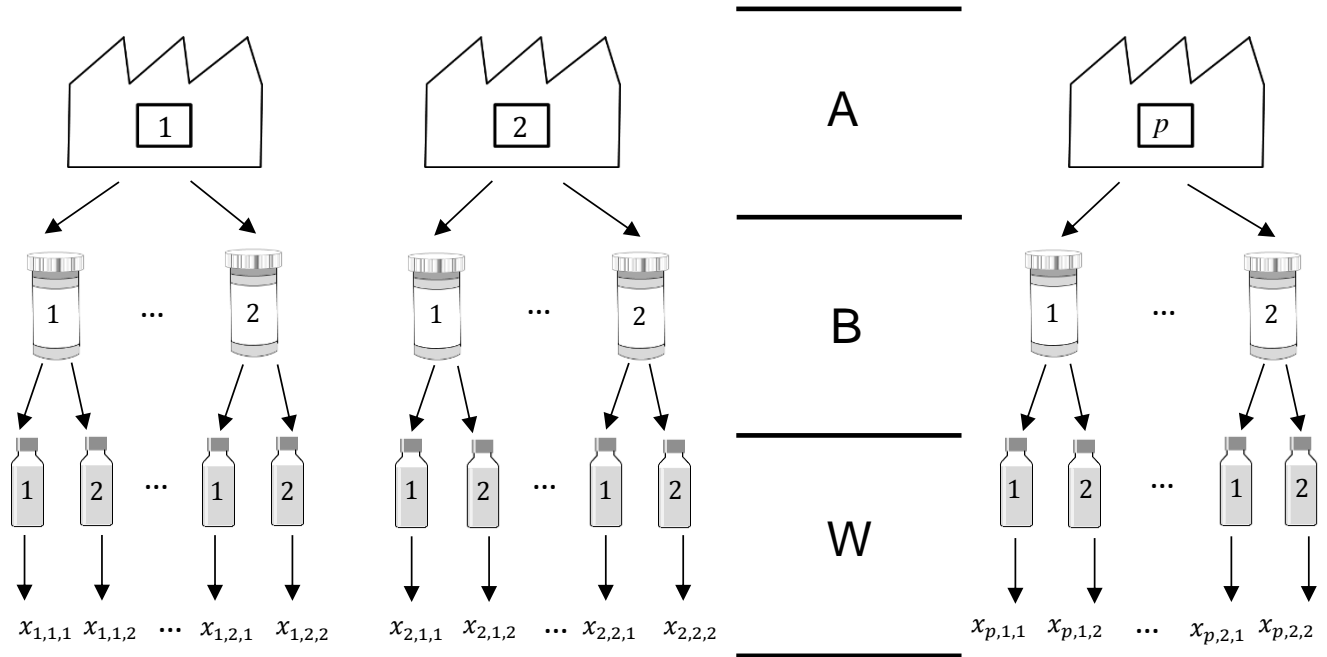
The  $A_i$  are assumed to be drawn from a population with mean 0 and standard deviation  $\sigma_L$ ,  $B_{ij}$  from a population with mean 0 and standard deviation  $\sigma_{\text{bb}}$  and  $\varepsilon_{ijk}$  from a population with mean zero and standard deviation  $\sigma_r$ . As before, the measurement error term is assumed to have the same variance, not only within each RM unit but, also within each laboratory. It is usually also assumed that the populations are normally distributed. The parameters to be obtained are the corresponding estimated between-laboratory standard deviation  $s_L$ , between-unit standard deviation  $s_{\text{bb}}$ , and repeatability standard deviation  $s_r$ . They are related as follows to the error terms:

$$\begin{aligned} s_L &= \sqrt{\text{Var}(A_i)} \\ s_{\text{bb}} &= \sqrt{\text{Var}(B_{ij})} \\ s_r &= \sqrt{\text{Var}(\varepsilon_{ijk})} \end{aligned} \quad (\text{B.7})$$

For the between-unit homogeneity ( $s_{\text{bb}}$ ), the same considerations apply with respect to the inability to detect batch heterogeneity as for the homogeneity study itself (see 7.8).

All these parameters can be estimated simultaneously by the analysis of variance (ANOVA) method,<sup>[56]</sup> if there are sufficient results of equal replication (the same number of replicate determinations from each unit and the same number of units per method/laboratory) after any technically or statistically invalid results have been excluded. If this ANOVA requirement cannot be met because of the number of invalid and/or missing results, the significance of the between-unit variance should be determined by other means (see Clause 7).

Theoretical details and additional methods for balanced and unbalanced ANOVA are given in standard textbooks (see, for example <sup>[58]</sup> <sup>[59]</sup> <sup>[60]</sup> <sup>[61]</sup>). A discussion of ANOVA in the context of the certification of reference materials is given in the literature (see, for example <sup>[16]</sup> <sup>[56]</sup> <sup>[62]</sup> <sup>[63]</sup>).



### Key

- A between-laboratory variation
- B between-unit variation (large containers indicate RM units)
- W measurement repeatability (small containers indicate individual extracts)

**Figure B.1 — Outline of a collaborative study, combined with batch homogeneity study**  
[characterization of an RM (2-way layout)]

The formulae for computing  $s_L$ ,  $s_{bb}$ , and  $s_r$  are as follows (see References <sup>[16]</sup> <sup>[13]</sup> <sup>[64]</sup>). The grand mean  $\bar{x}$  is computed using

$$\bar{x} = \frac{1}{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij}} \sum_{i=1}^p \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} x_{ijk} \quad (\text{B.8})$$

where  $p$  denotes the number of laboratories,  $b_i$  the number of units used by method/laboratory  $i$ , and  $n_{ij}$  is the number of replicate measurements on unit  $ij$ .  $s_L$ ,  $s_{bb}$ , and  $s_r$  are given by

$$s_r^2 = M_{\text{within}} \quad (\text{B.9})$$

$$s_{bb}^2 = \frac{M_B - M_{\text{within}}}{n_0} \quad (\text{B.10})$$

$$s_L = \frac{M_L - n_0' s_{bb}^2 - M_{\text{within}}}{n_0} \quad (\text{B.11})$$

The mean squares ( $M$ ) should be obtained using published methods (for example, References <sup>[16]</sup> and <sup>[64]</sup>) or a suitable statistical software package. The expressions given account for missing and/or removed (invalid) data. For complete data sets, the simpler formulae of ISO 5725-3 <sup>[14]</sup> may be used.

Following computation,  $s_{bb}$ , as calculated here may be used for the assessment of uncertainty associated with heterogeneity in the same way as that calculated in Clause 7.

When applied in an interlaboratory context, it is important to check the assumption of constant within-group variance. This assumption can be checked using statistical tests such as Bartlett's or Levene's tests.<sup>[58] [64]</sup>

### B.3 Linear regression (univariate linear model)

#### B.3.1 The basic model

Linear regression is used in reference material studies to determine simple rates of change and test their statistical significance. This clause describes the simplest case.

The simple linear model<sup>[67]</sup> can be expressed as

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (\text{B.12})$$

where  $\beta_0$  and  $\beta_1$  are the intercept and slope, respectively and  $\varepsilon$  denotes the random error component, usually assumed to be normally distributed with a mean of zero.

Given a set of  $n$  pair-wise observations of  $Y$  versus  $X$ , individual observations  $(x_i, y_i)$  are related by

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (\text{B.13})$$

**EXAMPLE** For a basic stability study,  $x_i$  corresponds to an observed time and  $y_i$  to the corresponding observed value of the property of interest.

#### B.3.1 Fitting the model

The regression parameters can be computed from the following expressions:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{B.14})$$

$$b_0 = \bar{y} - b_1 \bar{x} \quad (\text{B.15})$$

where  $b_0$  and  $b_1$  are the estimated intercept and slope, respectively and  $\bar{x}$  and  $\bar{y}$  are the mean of the respective observations.

The standard errors  $s(b_1)$  and  $s(b_0)$  in  $b_1$  and  $b_0$  can be computed as follows:

$$s(b_1) = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (\text{B.16})$$

where

$$s^2 = \frac{\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2}{n - 2} \quad (\text{B.17})$$

and

$$s(b_0) = s(b_1) \sqrt{\frac{\sum_{i=1}^n x^2}{n}} \quad (\text{B.18})$$

### B.3.3 Inspection and check of assumptions

Typical review of the preliminary fitted model includes:

- calculation of residuals [ $r_i = y_i - (b_0 + b_1 x_i)$ ] and plotting against time, followed by inspection for evidence of curvature, run-to-run effects, substantial differences in dispersion from one time point to another, or for outlying values;
- preparation of a normality (or “Q-Q”) plot of residuals and inspection for evidence of non-normality;
- optionally, tests for significant differences in dispersion, significant between-group differences in residuals, or for nonlinearity.

### B.3.4 Testing for statistically significant change

For the simple case above, after confirming the validity of the basic regression assumptions, the usual test for a statistically significant gradient is a  $t$ -test for slope significantly different from zero. This is conducted by calculating the  $t$  statistic

$$t_{b_1} = \frac{|b_1|}{s(b_1)} \quad (\text{B.19})$$

and comparing this with the two-tailed critical value of Student's  $t$  for  $n - 2$  degrees of freedom at the 95 % level of confidence. If the calculated test statistic  $t_{b_1}$  exceeds the critical value, the slope is considered to be significantly different from zero at the 95 % level of confidence.

### B.3.5 Confidence interval for the regression line

The fitted values  $b_0$  and  $b_1$  may be used to provide an estimate  $\hat{y}$ . of the value of the response for a particular value  $\hat{x}$ , using

$$\hat{y} = b_0 + b_1 \hat{x} \quad (\text{B.20})$$

It is sometimes useful to obtain a confidence interval for the predicted value  $\hat{y}$ . A two-sided confidence interval at the 95 % level of confidence is given by

$$\hat{y} \pm t_{95, n-2} s \sqrt{\frac{1}{n} + \frac{(\hat{x} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (\text{B.21})$$

where  $t_{95, n-2}$  is the two-tailed critical value for Student's  $t$  with  $n - 2$  degrees of freedom. Confidence intervals for different levels of confidence may be similarly obtained by using the appropriate critical value for Student's  $t$ .

NOTE This interval is not the prediction interval for individual values  $y_i$ .

## B.4 Predicting shelf life or choosing initial monitoring point based on stability study results

### B.4.1 Principle

Where there is an acceptable range, described by upper and/or lower specification limits, for change in a property value over time in storage, it can be useful to estimate either the storage lifetime (or ‘shelf life’) of a reference material, or set a first monitoring point, using the results from a stability study.

The principle is to estimate the 95 % confidence interval for future values, taking account of the estimated degradation rate, and to choose the shortest time at which one of the confidence limits (upper or lower) intersects a specification limit. This principle is illustrated schematically in Figure B.2.

For a CRM, any acceptable range chosen for the certified value needs to be less than the expanded uncertainty, ideally less than  $U_{CRM}/3$ . For non-certified reference materials, any acceptable change needs to be sufficiently small to avoid adverse effects on the intended use.

#### B.4.2 Prediction of shelf life in the case of a linear trend

Let the upper and lower limits of an acceptable range of values for a RM property be  $L_{\text{upr}}$  and  $L_{\text{lwr}}$ , respectively. The confidence interval around a predicted future value of an RM when the degradation follows a simple linear form is given by Formulae (B.20) and (B.21), taking  $x$  as the time from the beginning of the stability study. The required shelf life is obtained by setting Formula (B.21) to each of the specification limits and solving for  $\hat{x}$  to obtain the shortest positive shelf life  $t_{\text{shelf}}$ . That is,  $t_{\text{shelf}}$  is chosen according to Formula (B.22):

$$t_{\text{shelf}} = \min(t_{s,\text{upr}}, t_{s,\text{lwr}}) \quad (\text{B.22})$$

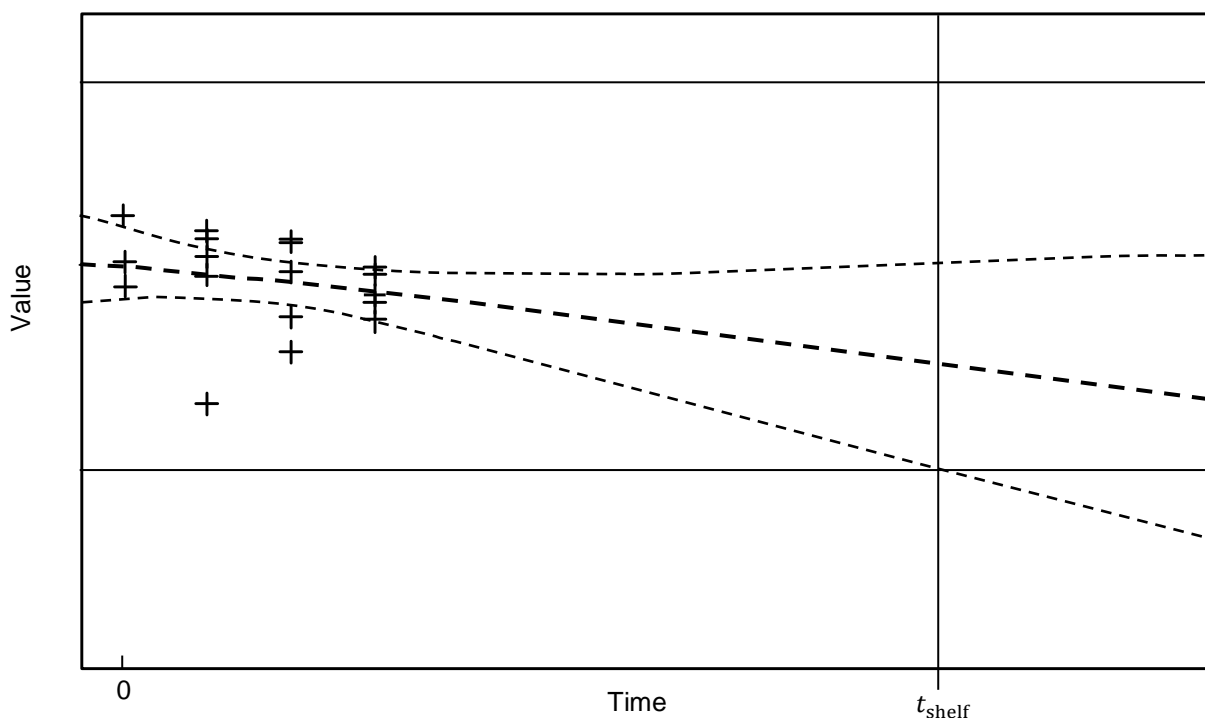
where  $t_{s,\text{upr}}$  and  $t_{s,\text{lwr}}$  are the lowest positive solutions to

$$L_{\text{upr}} - \left[ b_0 + b_1 t_{s,\text{upr}} \pm t_{95,n-2^S} \sqrt{\frac{1}{n} + \frac{(t_{s,\text{upr}} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right] = 0 \quad (\text{B.23})$$

$$L_{\text{lwr}} - \left[ b_0 + b_1 t_{s,\text{lwr}} \pm t_{95,n-2^S} \sqrt{\frac{1}{n} + \frac{(t_{s,\text{lwr}} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right] = 0 \quad (\text{B.24})$$

NOTE 1 Formulae (B.23) and (B.24) have, in total, up to four real solutions. Where there is no positive solution for one of these equations – for example for statistically significant degradation in which both positive solutions arise from Formula (B.24) – the negative solutions are disregarded.

NOTE 2 Formulae (B.23) and (B.24) cannot be solved algebraically; both must be solved iteratively, for example using a search method.



NOTE Crosses show data points for the stability study starting at time zero. Solid horizontal lines show upper and lower specification limits. The straight dashed line is the least squares line fitted using the procedure of Clause B.3. Curved dashed lines are the upper and lower confidence intervals for the line calculated using Formula (B.21). The solid vertical line is at  $t_{shelf}$ .

**Figure B.2 — Principle of shelf life calculation from a stability study**

NOTE 3 If the specification limits are too close together [specifically, if either specification limit is inside the interval given by Formula (B.21)], it is not possible to find a real-value solution.

### B.4.3 Prediction of shelf life for other degradation models

For different degradation models, shelf life can be predicted following the same general principles as described in B.4.2, with the confidence interval calculation, and consequently Formulae (B.23) and (B.24), adjusted to suit the particular degradation model.

### B.4.4 Choosing an initial monitoring point

The same principle may be used to choose an initial monitoring point for post-certification monitoring of stability; an acceptable change in property value is chosen and the value of  $t_{shelf}$  calculated as above then indicates the time from the beginning of the stability study that monitoring should first take place.

## B.5 Robust statistics

### B.5.1 Robust estimators for interlaboratory studies

A range of basic robust estimators suitable for interlaboratory studies is given in Table B.1, together with literature sources and some key performance indicators. Their use is subject to the following.

- g) Each laboratory's data should be represented by a single value (which may itself be a mean, median or other location estimate);
- h) Information and results from the laboratories should be scrutinised for evidence of technical errors and results found to be technically invalid should be removed from the data set before application of robust methods;
- i) There should not normally be less than 10 data points after any results have been removed for technical reasons;
- j) The majority of values in the data set (usually the central portion) should be approximately normally distributed or at least symmetrically distributed (after transformation, if required);
- k) There should be no significant grouping by test method, geographical origin, etc.;

- l) The proportion of extreme outlying values in the data set should not exceed half of the breakdown point of the estimator in use.

The robust estimators shown in Table B.1 do not form an exclusive list; Algorithm A and the Hampel estimator may both be tuned by changing the tuning parameters and a wide range of alternative algorithms exists. Variations on the algorithms shown, and other robust estimators, may be used subject to demonstration of similar or better performance than those in Table B.1. Performance of other estimators may be assessed from the theoretical breakdown point, the relative efficiency and any additional properties (such as resistance to asymmetry or performance on smaller data sets) that may be appropriate for the application.

The number of data points at c) above may be below 10 if the RM producer can demonstrate adequate performance of the chosen estimator on typical data sets for the application in question.

NOTE 1 The upper limit on proportion of extreme values is intentionally well below the breakdown point of the estimators both to avoid adverse effects of an increasing proportion of outliers and because above these proportions it is normally appropriate to consider whether the certification exercise should be abandoned for lack of agreement between laboratories.

NOTE 2 The estimators shown in Table B.1 do not use reported uncertainties or information on within-laboratory precision. They, and the simple arithmetic mean, are therefore most appropriate when uncertainties associated with the laboratory mean values are either not available or not all reliable.

NOTE 3 Outliers may adversely affect some statistical procedures (including analysis of variance) that can be used to test for significant grouping effects. For this purpose, it can be useful to repeat the statistical test after removal of extreme outliers in order to assess the effect of outliers on the statistical test used.

NOTE 4 Most robust estimators can be tuned for increased resistance to outliers at the expense of decreased relative efficiency by changing tuning parameters. The best balance between efficiency (which affects uncertainty) and outlier resistance depends on the distribution of the data, including the outliers. Tuning parameters that give relative efficiency between 0,80 and 0,95 for a Normal distribution are typical.

**Table B.1 — Some robust estimators and their properties**

Robust method	Population parameter estimated (Note 1)	Relative efficiency $e_R$ (Note 2, Note 3)	Breakdown Point (Note 4, Note 5)	Reference (Note 6)
Sample mean	Mean	1,0	0 %	
Median	Mean	0,67	50 %	
Algorithm A	Mean and standard deviation	0,97 <sup>a,b</sup> , 0,73 <sup>c</sup>	25 %	ISO 5725-3, ISO 13528
Hampel (with $Q_n$ or MADe scale)	Mean	0,96 <sup>d</sup>	50 %	ISO 13528
Sample standard deviation	Standard deviation	1,0	0 %	
Normalized interquartile range (nIQR)	Standard deviation	0,37	25 %	ISO 13528
Scaled median absolute deviation (MADe)	Standard deviation	0,37	50 %	ISO 13528
$Q_n$	Standard deviation	0,81	50 %	ISO 13528
NOTE 1 Strictly, the population referred to here is the underlying normal distribution that is assumed to have been contaminated with outliers.				
NOTE 2 Relative efficiency is a measure of the variability of the estimator when applied to normally distributed data. It is the variance of the estimator divided by the variance of the arithmetic mean or (for estimators of standard deviation) by the sample standard deviation on the same data. High efficiency is better. The efficiencies shown are large-sample efficiency; for the estimators shown here the relative efficiency (compared to the mean or standard deviation) improves slightly for smaller data sets.				



NOTE 3 While robust estimators are not as efficient as the mean for normally distributed data (relative efficiency < 1,0), for other distributions, with heavier tails or outliers, robust estimators quickly improve over the arithmetic mean; the same is true for the sample standard deviation and the corresponding robust estimators of standard deviation.

NOTE 4 The breakdown point is a measure of the estimator's resistance to a proportion of extreme outliers. Higher is better.

NOTE 5 The definition of breakdown point used here is the proportion of a large normally distributed data set that can be moved to +infinity without the estimate also moving to infinity. For example, if just under 50 % of a data set is replaced by +infinity, the median will remain within the remaining finite data.

NOTE 6 References to ISO Standards have been given, other than for the mean, median and sample standard deviation, which are well known.

NOTE 7 Performance information in this table is taken from ISO 13528:2015.<sup>[10]</sup>

- a For the tuning parameter,  $k = 1,5$
- b Efficiency for the Algorithm A estimate of the mean with tuning parameter  $k = 1,5$  as in ISO 5725-3.
- c Efficiency for the Algorithm A estimate of the standard deviation with tuning parameter  $k = 1,5$
- d Efficiency for the Hampel algorithm with tuning parameters  $a = 1,5$ ,  $b = 3,0$ ,  $c = 4,0$  as in ISO 13528.

### B.5.2 Estimation of uncertainties for robust estimators

The uncertainty for any estimate based on a location estimate derived directly from interlaboratory data depends in part on the estimator used and in part on the distribution from which the data are drawn. For normally distributed data, the statistical uncertainty associated with a robust estimate is

$$\frac{S^*}{\sqrt{e_R N}} \quad (\text{B.25})$$

Robust estimates are, however, rarely used on normally distributed data; they are used on data that includes a higher proportion of values far from the (true) mean and that may include very extreme outlying results. Under these circumstances, the uncertainty is always larger than that given by Formula (B.25). The most appropriate uncertainty then depends strongly on the distribution. The most general method of assigning uncertainties for estimates from appreciably non-normal data are to use simulation or re-sampling methods. In the absence of uncertainties based on simulation studies or resampling from the particular data set, the characterization uncertainty for robust estimates should not be less than

$$1,253 \frac{S^*}{\sqrt{N}} \quad (\text{B.26})$$

The factor 1,253 is based on the standard deviation of the median, or the efficiency of the median as an estimate of the mean, in a large set of results drawn from a normal distribution. While the efficiency of more sophisticated robust methods can be much greater than that of the median, this factor has been recommended as a conservative (high) estimate, to allow for modest non-normality.

## Příloha C (informativní)

### Příklady

#### C.1 Studie homogenity

Byla provedena studie homogenity mezi jednotkami v projektu pro přípravu certifikovaného referenčního materiálu pro chrom v půdě. Výsledky této studie homogenity udává Tabulka C.1

**Tabulka C.1 — Výsledky studie homogenity mezi jednotkami pro stanovení chromu v půdě**

Lahvička č.	Výsledek č. 1	Výsledek č. 2	Výsledek č. 3
1	121,30	128,74	119,91
2	120,87	121,32	119,24
3	122,44	122,96	123,45
4	117,60	119,66	118,96
5	110,65	112,34	110,29
6	117,29	120,79	121,42
7	115,27	121,45	117,48
8	118,96	123,78	123,29
9	118,67	116,67	114,58
10	126,24	123,51	126,20
11	128,65	122,02	121,93
12	126,84	124,72	123,14
13	122,61	128,48	126,20
14	118,95	123,82	118,11
15	118,74	118,23	117,38
16	119,74	121,78	121,01
17	121,21	123,28	116,38
18	129,30	124,10	122,02
19	136,81	129,80	128,47
20	127,81	117,66	122,90
POZNÁMKA Všechny výsledky jsou v mg/kg			

Analýza rozptylu uvedená v Tabulce C.2 se vypočítá z údajů v Tabulce C.1 např. pomocí počítačového softwaru (tabulkového procesoru).

**Tabulka C.2 — Tabulka ANOVA pro studii homogenity mezi lahvičkami pro stanovení chromu v půdě**

Zdroj variability	Suma čtverců	Stupně volnosti	Průměrný čtverec <i>M</i>
Mezi lahvičkami	1 037,1	19	54,59
Uvnitř lahviček	330,5	40	8,26
Celkem	1 367,6	59	

Rozptyl mezi skupinami (rozptyl mezi jednotkami v této studii homogenity) se odhadne pomocí vztahu

$$s_{\text{between}}^2 = \frac{M_{\text{between}} - M_{\text{within}}}{n_0} = \frac{54,59 - 8,26}{3} \text{mg}^2\text{kg}^{-2}$$

Směrodatná odchylka mezi jednotkami je druhou odmocninou tohoto rozptylu

$$s_{bb} = \sqrt{15,44} = 3,93 \text{ mg kg}^{-1}$$

Směrodatná odchylka opakovatelnosti se vypočítá z  $M_{\text{within}}$

$$s_r = \sqrt{M_{\text{within}}} = \sqrt{8,26} = 2,87 \text{ mg kg}^{-1}$$

## C.2 Homogenita mezi jednotkami pro případ omezené opakovatelnosti testovací metody

Příklad je převzat z Reference [65], certifikace katalytické koncentrace gammaglutamyltransferázy ve tkáni ledvin prasete. Tkáň prasečích ledvin byla zředěna chloridem sodným a homogenizována. Po opakovaném vysrážení byl produkt podroben chromatografickému přečištění na koloně DEAE Trisacryl a poté proběhla chromatografie na hydroxyapatitové koloně. Jako materiál byl vybrán hovězí sérový albumin o koncentraci 60 g/l, protože jeho přísávek nemodifikoval katalytické vlastnosti částečně vyčištěného enzymu. Proces rozplňování byl kontrolován vážením 101 lahvíček odebíraných v pravidelných intervalech. Nebyl zjištěn žádný trend během rozplňování. Nakonec byl materiál lyofilizován.

Materiál byl kontrolován na obsah kontaminujících enzymů, čistotu, obsah vody a zbytkového kyslíku a byl shledán vhodným jako CRM. Pro výpočet variability mezi jednotkami byla vzata data z posuzování homogenity z dřívější certifikační studie, která zahrnovala 20 ampulí, každá měřená 6x. Byla provedena analýza rozptylu (ANOVA) pro skupiny ampulí. Byla vypočtena směrodatná odchylka ( $s_r$ ) uvnitř lahvíček a mezi lahvíčkami ( $s_{bb}$ ). Výsledky analýzy rozptylu včetně průměru čtverců a odhadu směrodatných odchylek na různých úrovních jsou uvedeny v Tabulce C.3.

**Tabulka C.3 — Výsledky katalytických koncentrací gamma-glutamyl transferázy ve tkáni prasečích ledvin (relativní hodnoty v závorkách)**

Průměr	67,78 IU/l
$M_{\text{between}}$	1,76 IU <sup>2</sup> /l <sup>2</sup>
$M_{\text{within}}$	1,63 IU <sup>2</sup> /l <sup>2</sup>
$s_r$	1,28 IU/l (1,88 %)
$s_{bb}$	0,147 IU/l (0,22 %)
$u'_{bb}$	0,196 IU/l (0,29 %)

Směrodatná odchylka mezi jednotkami se vypočítá pomocí vztahu

$$s_{bb}^2 = \max\left(\frac{M_{\text{between}} - M_{\text{within}}}{n_0}, 0\right) = \frac{1,76 - 1,63}{6} = 0,0217$$

který vede k odhadu pro  $s_{bb}$  0,147 IU/l, který je ekvivalentní hodnotě 0,22 % (rel.). V tomto případě, směrodatná odchylka uvnitř skupin má 100 stupňů volnosti. Směrodatná odchylka opakovatelnosti se vypočítá z  $M_{\text{within}}$

$$s_r = \sqrt{M_{\text{within}}} = \sqrt{1,63} = 1,28 \text{ IU/l}$$

Tato hodnota byla považována za příliš velkou, aby poskytovala správný důkaz o heterogenitě tak nízké jako 0,22 %, rel. Výrobce si proto vybral konzervativnější hodnotu pro nejistotu přidruženou k heterogenitě, označenou  $u'_{bb}$ , založenou na doporučení v referenci [66] a to následovně pomocí vztahu

$$u'_{bb} = \sqrt{\frac{M_{\text{within}}}{n_0}} \cdot \sqrt{\frac{2}{v M_{\text{within}}}} = \sqrt{\frac{1,63}{6}} \sqrt{\frac{2}{100}} = 0,196 \text{ IU/l}$$

(Tento výraz ukazuje, že je odhadem standardní nejistoty pro směrodatnou odchylku uvnitř jednotek).

Reference [16] doporučuje použití vyšší z hodnot  $s_{bb}$  a  $u'_{bb}$  uvedených shora; hodnota 0,196 IU/l byla proto použita jako nejistota přidružená k rozptylu mezi jednotkami.

### C.3 Hodnocení homogenity při použití znáhodněného blokového návrhu

#### C.3.1 Podklady a data

Byla provedena studie homogenity mezi jednotkami v projektu pro přípravu referenčního materiálu certifikovaného na obsah aniontů v říční vodě. Tento příklad používá data získaná pro sírany. Studie byla provedena formou znáhodněného blokového návrhu, ve kterém bylo 15 jednotek RM měřeno jednou, každá ve třech různých měřících cyklech (sériích). Každý cyklus byl náhodně sestaven samostatně, tj. jednotlivé jednotky byly měřeny v každém cyklu v různém pořadí. Data pro sírany, seřazená podle čísel jednotek RM, jsou uvedena v Tabulce C.4.

**Tabulka C.4 — Údaje pro homogenitu obsahu síranů (mg.l<sup>-1</sup>)**

Jednotka	Cyklus 1	Cyklus 2	Cyklus 3
15	79,737	79,036	80,196
127	80,177	78,367	80,005
199	79,950	78,555	80,040
248	78,970	79,070	79,299
309	79,164	78,865	79,567
375	79,315	77,810	80,250
439	79,720	79,142	80,424
482	80,102	78,217	80,342
571	80,457	78,569	80,022
662	80,105	78,280	80,770
689	80,550	79,333	80,530
775*	79,215	78,442	77,919
841	79,96	78,947	80,313
911	79,637	79,574	80,107
973	79,710	78,614	80,385
Pořadí jednotlivých cyklů, které bylo znáhodněno, není uvedeno.			
* Jednotka 775 vykazuje odlehlou hodnotu a byla vyřazena; viz text.			

Vynesení všech dat z Tabulky C.4 do grafu (Obr.C.1) ukazuje jasný vliv jednotlivých cyklů a jednu odlehlou hodnotu (identifikovanou jako hodnotu jednotky 775) v cyklu 3. Bylo vysledováno, že tato hodnota patří špatnému tvaru chromatografického píku. Prostředky pro odstranění odlehlé hodnoty zahrnují odstranění individuálního bodu nebo odstranění dotčené jednotky RM. V tomto případě, aby se zachovalo vyvážené uspořádání projektu a pro zjednodušení analýzy dat, byla všechna tři opakovaná měření pro tuto jednotku odstraněna, ponechávající tak technicky platná data pro 14 jednotek.

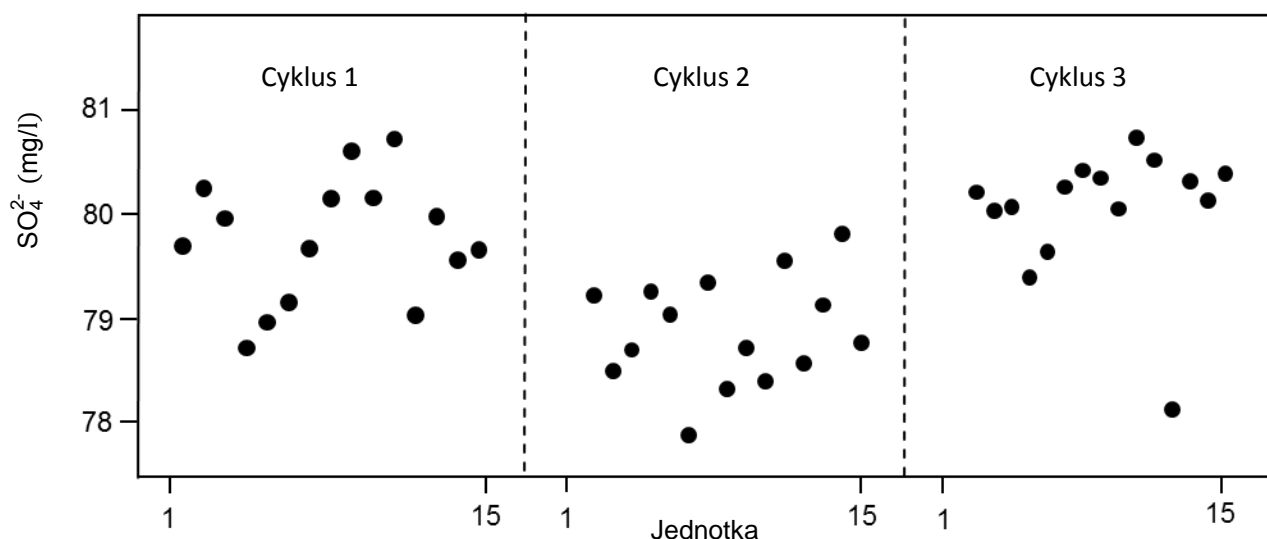
#### C.3.2 Vyhodnocení

Dvoufaktorová (často také nazývaná „dvoucestná“) analýza rozptylu (ANOVA), bez opakování měření byla použita při zpracování výsledků z Tabulky C.4. Výsledná ANOVA, uvedená v Tabulce 5, potvrzuje významný vliv cyklu.

**Tabulka C.5 — Tabulka dvoufaktorové ANOVA pro sírany**

	Df <sup>a</sup>	Suma Sq <sup>b</sup>	Průměr Sq <sup>c</sup>	F hodnota <sup>d</sup>	Pr(> F) <sup>e</sup>
Jednotka	13	3,065 1	0,2358	1,3476	0,249 6
Cyklus	2	15,409 3	7,704 6	44,036 4	4,48e-09
Rezidua	26	4,549 0	0,175 0		

- |   |  |
|---|--|
| a | počet stupňů volnosti  |
| b | suma čtverců   |
| c | průměrný čtverec   |
| d | F- statistika (průměrný čtverec dělený reziduálním průměrným čtvercem)   |
| e | hodnota $p$ (pravděpodobnost překročení pozorované hodnoty F-statistiky při nulové hypotéze nulového vlivu rozptylu) |



POZNÁMKA Svislé čárkované čáry oddělují jednotlivé cykly. Čísla jednotek 1 až 15 v tomto grafu odpovídají číslům 15 jednotek v Tabulce C.4, v pořadí čísel jednotek.

**Obrázek C.1 — Naměřené hodnoty síranů v kandidátském referenčním materiálu**

Rozptyly uvnitř a mezi jednotkami  $s_w^2$  a  $s_{bb}^2$  jsou odhadnuty, jak je ukázáno v podkapitole 7.7.4.

$$s_w^2 = s_r^2 = M_w = 0,175 0$$

Kde  $M_w$  značí průměrný čtverec mezi jednotkami, poskytující  $s_w = 0,4183 \text{ mg l}^{-1}$  a

$$s_{bb}^2 = \max\left(\frac{M_A - M_w}{n_0}, 0\right) = \max\left(\frac{0,235 8 - 0,175 0}{3}, 0\right) = 0,020 3$$

kde  $M_A$  značí průměrný čtverec mezi jednotkami a  $n_0$  počet pozorování pro (každou) jednotku (identický s počtem cyklů v tomto jednoduchém znáhodněném blokovém návrhu). To dává  $s_{bb} = 0,1424 \text{ mg l}^{-1}$ .

POZNÁMKA Například, jednofaktorová analýza rozptylu, aplikovaná na stejná data seskupená pouze podle jednotek poskytuje stejný průměrný čtverec mezi jednotkami (0,2358) ale také značně „nafouklý“ residuální průměrný čtverec 0,7128. Toto by mohlo nesprávně vést k nulovému odhadu rozptylu mezi jednotkami  $s_{bb}$ , podhodnocující tak nejistotu přidruženou k homogenitě, nebo pokud je příspěvek homogenity založen na výrazu pro „uvnitř jednotek“ mohlo by to vést k podstatnému nadhodnocení.

## C.4 Hodnocení homogenity: jednoduchý návrh s korekcí lineárního trendu

### C.4.1 Podklady a data

Byla provedena studie homogenity mezi jednotkami v projektu na přípravu certifikovaného referenčního materiálu pro kongenery whisky. 35 jednotek RM bylo podrobena duplicitní analýze pro testování homogenity v rámci jedné chromatografické série, replicitní analýzy se uskutečnily v náhodném pořadí. Tabulka 10 udává výsledky pro tuto studii homogenity a residua pro lineární přiřazení k těmto datům (viz níže). V této tabulce, „číslo jednotky“ odpovídá pořadí, ve kterém byla naplněna, zatímco „Pozice“ znamená pořadí v měřicí sérii. Podle 7.7.1 byly

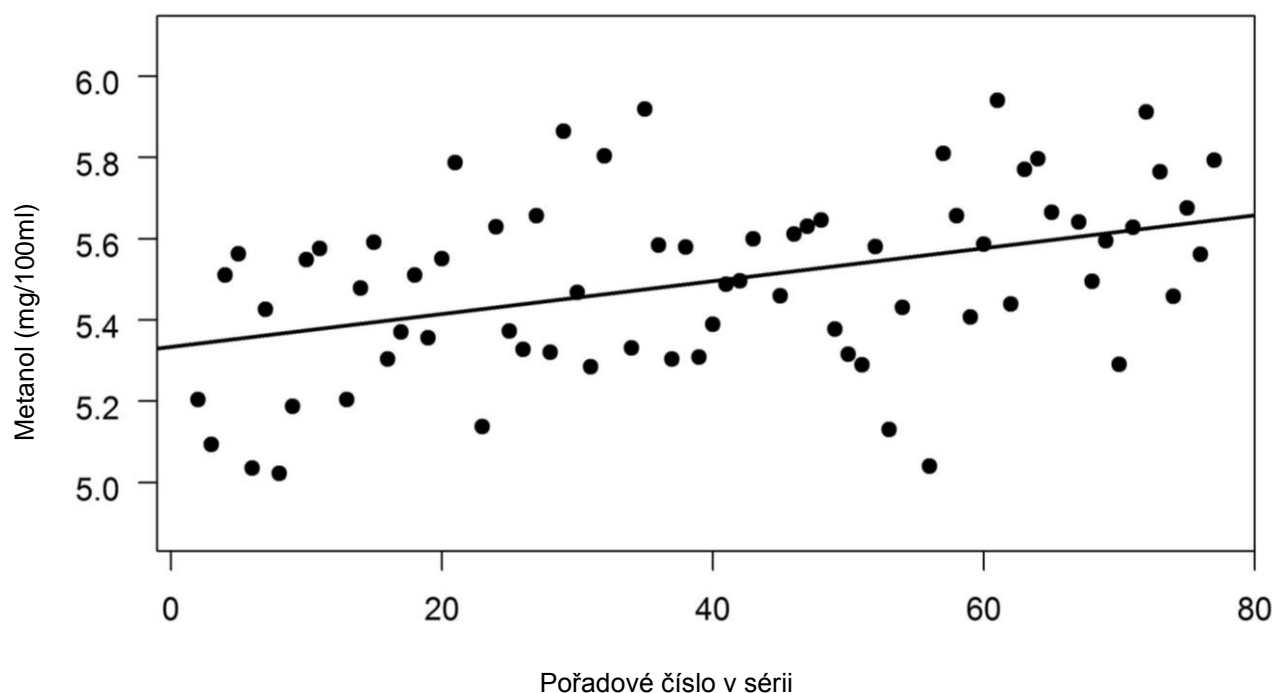
výsledky vyneseny do grafu „jednotka“ versus „Pořadí v sérii“. Graf je uveden na obrázku C.2. Mírný vzestupný trend, který lze přičíst přístrojovému driftu, je zřejmý; plná čára je přímka zkonstruovaná metodou nejmenších čtverců z pořizovaných dat. Výsledky regresní analýzy jsou uvedeny v Tabulce C.6. a ukazují velmi silně významný vliv „Pozice“ (směrnice) na který ukazuje nízká  $p$ -hodnota („ $Pr(> |t|)$ “ v tabulce).

**Tabulka C.6 — Regresní statistika pro metanol v závislosti na pořadí v měřicí sekvenci**

	Odhad <sup>a</sup>	Směrodatná chyba <sup>b</sup>	t hodnota <sup>b</sup>	(Pr(>  t )) <sup>c</sup>
Úsek	5,333 7	0,049 8	107,15	< 2e-16
Pozice	0,004 0	0,001 1	3,69	0,000 45

<sup>a</sup> Koefficienty a směrodatné chyby jsou uvedeny na čtyři desetinná místa  
<sup>b</sup>  $t$ -statistika (odhad dělený směrodatnou chybou)  
<sup>c</sup>  $p$ -hodnota (pravděpodobnost překročení sledované absolutní hodnoty  $t$ -statistiky při nulové hypotéze nulového gradientu)

Trend není tak výrazný, aby to vyžadovalo vynětí (odmítnutí) celé série a opětovnou analýzu, a vizuální kontrola grafu ukazuje, že lineární průběh je přiměřenou aproximací. Proto se přistoupilo ke statistickému zpracování s cílem odstranit nepříznivý vliv trendu. Jsou ilustrovány dvě možnosti: korekce následovaná klasickou analýzou rozptylu a (pouze pro porovnání výsledků) výsledky smíšeného vlivu modelování, které zahrnují současně vliv trendu a vliv každé jednotky.



POZNÁMKA Plná čára odpovídá lineárnímu proložení údaji. Číslo pořadí v sérii odpovídá označení „Pozice“ v tabulce C.10.

**Obrázek C.2 — Výsledky stanovení metanolu v kandidátském RM whisky vynesené v pořadí měření**

#### C.4.2 Metoda 1 Korekce trendu následovaná klasickou analýzou rozptylu (ANOVA)

Počáteční krok odečítá z dat lineární trend. Výsledky jsou uvedeny jako „rezidua“ v Tabulce C.10.

POZNÁMKA Reziduály odpovídají odečtení jak příspěvku pro úsek, tak pro polohu v lineárním modelu. Protože následná analýza rozptylu předpokládá aditivitu jak vlivu jednotky, tak vlivu náhodné chyby, bylo by stejně platné odečíst člen pro pozici a ponechat data bez korekce na úsek.

Jednofaktorová analýza rozptylu (ANOVA) je poté aplikována na reziduály s „jednotkou“ jako faktorem seskupení. Počáteční tabulka ANOVA s daty korigovanými na trend je uvedena v Tabulce C.7.

**Tabulka C.7 — Počáteční ANOVA aplikovaná na reziduály homogenity whisky**

	Df <sup>a</sup>	Suma Sq <sup>b</sup>	Průměr. Sq <sup>c</sup>	F hodnota <sup>d</sup>	(Pr(> F)) <sup>e</sup>
Mezi jednotkami	34	1,764 6	0,051 90	1,714 6	0,058 65
Uvnitř jednotek	35	1,059 5	0,030 27	0,000 45	
<p><sup>a</sup> počet stupňů volnosti</p> <p><sup>b</sup> suma čtverců</p> <p><sup>c</sup> průměrný čtverec</p> <p><sup>d</sup> F- statistika (průměrný čtverec dělený reziduálním průměrným čtvercem)</p> <p><sup>e</sup> hodnota <math>p</math> (pravděpodobnost překročení pozorované hodnoty F-statistiky při nulové hypotéze nulového vlivu rozptylu)</p>					

Protože reziduály v Tabulce C10 vznikají z lineárního modelu s výrazy pro gradient (Pozice) a úsek, je však tabulka pro ANOVA uvedená v Tabulce C.7 mírně nepřesná, protože ANOVA předpokládá, že se odečítá pouze celkový průměr. Provádí se tudíž malá korekce, provedená pomocí

- odečtením dalších stupňů volnosti od stupňů volnosti „uvnitř“ jednotek v tabulce ANOVA při výpočtu výrazu pro gradient a
- přepočítáním reziduálního průměrného čtverce dělením příslušné sumy čtverců korigovaných počtem stupňů volnosti. Výsledná korigovaná tabulka ANOVA je uvedena (s odpovídajícími výpočty) jako Tabulka C.8.

Zatímco korekce průměrného čtverce uvnitř jednotek je malá „mezi jednotkový“ výraz používá rozdíl mezi průměrnými čtverci a korekce může být někdy větší než požadovaný rozdíl.

Jestliže je pro korekci trendu použit složitější model, korekce stupňů volnosti uvnitř jednotek by měla respektovat počet parametrů použitých v modelu. Typicky, jestliže existuje v modelu  $q$  parametrů,  $q-1$  se odečítá od počtu stupňů volnosti uvnitř skupin.

**Tabulka C.8 — Tabulka pro korigovanou ANOVA lineárním modelem stupňů volnosti**

	Df <sup>a</sup>	Suma Sq <sup>b</sup>	Průměr. Sq <sup>c</sup>	F hodnota <sup>d</sup>	(Pr(> F)) <sup>e</sup>
Mezi jednotkami	34	1,764 6	0,051 90	1,714 6	0,058 65
Uvnitř jednotek	(35–1) = 34	1,059 5	1,059 5/34 = 0,031 16		
<p><sup>a</sup> počet stupňů volnosti</p> <p><sup>b</sup> suma čtverců</p> <p><sup>c</sup> průměrný čtverec</p> <p><sup>d</sup> F- statistika (průměrný čtverec dělený reziduálním průměrným čtvercem)</p> <p><sup>e</sup> hodnota <math>p</math> (pravděpodobnost překročení pozorované hodnoty F-statistiky při nulové hypotéze nulového vlivu rozptylu)</p>					

Stanovení směrodatné odchylky uvnitř a mezi jednotkami  $s_{bb}$  a  $s_w$  nyní probíhá jako v podkapitole 7.7.3

$$s_{bb}^2 = \max\left(\frac{MS_b - MS_w}{n_0}, 0\right) = \max\left(\frac{0,051\ 90 - 0,031\ 16}{2}, 0\right) = 0,010\ 4\ \text{mg}/100\ \text{ml}$$

a poskytuje směrodatnou odchylku mezi jednotkami  $s_{bb}$  0,108 mg/100ml. Směrodatná odchylka uvnitř jednotek  $s_w$  je druhou odmocninou reziduálního průměrného čtverce, tj. 0,1765 mg/100 ml.

Pro srovnání, odhady  $s_{bb}$  a  $s_w$  získané z analýzy dat bez korekce trendu byly v tomto případě 0,0928 mg/100 ml a 0,2016 mg/100 ml, mírně podhodnocující směrodatnou odchylku mezi jednotkami a nadhodnocující směrodatnou odchylku uvnitř jednotek.

### C.4.3 Metoda 2: Zahrnutí trendu v cyklu do analýzy

Ve statistické terminologii, vliv samotného trendu v cyklu je nejběžněji modelován jako tzv. „fixní vliv“, zatímco vliv mezi jednotkami se považuje za náhodný. Nejobecnější metodou jak zahrnout a zjistit stálé a náhodné vlivy společně je aplikace tzv. modelování smíšených vlivů (viz. např. Reference [26]). Protože předmětem našeho zájmu jsou minimálně vychýlené rozptyly, uvnitř jednotek a mezi jednotkami, vhodným kritériem fitování je omezená maximální věrohodnost (REML) (viz 7.7.6). To vyžaduje vhodný statistický software. V následujícím příkladu byl použit nlme-sofwarový balík[67] z balíku R otevřeného software[68]. Protože specifikace modelu a jeho provedení zcela závisí na použitém softwaru, jsou předmětem zájmu pouze výsledky. Ty jsou uvedeny, spolu s krátkým komentářem, ke každé části, v tabulce C.9.

Povšimněte si, že výsledky těsně souhlasí s klasickými výpočty uvedenými v kapitole C.2. Hlavními výhodami jsou:

- Je-li k dispozici vhodný software, výpočet vyžaduje méně operací. V tomto případě, výsledky uvedené v Tabulce C.9 vyžadují jednoduchý příkaz a žádné další výpočty. To je v protikladu s odděleným modelem lineárního fitování, výpočtem tabulky ANOVA, korigovaným výpočtem tabulky ANOVA a manuálním výpočtem směrodatných odchylek tak, jak to požaduje klasický přístup, použitý v kapitole C.4.2.
- Protože model současně v sobě zahrnuje jak náhodné vlivy mezi jednotkami, tak trend, lineární model lépe zapracovává jakékoliv částečné spletení vlivu pozice jednotky a pořadí série měření, které mohou příležitostně vést v nadhodnocenou nebo podhodnocenou interpretaci výsledků lineární regrese aplikované bez zahrnutí náhodné chyby v důsledku vlivů mezi jednotkami.
- Používání REML kritéria poskytuje minimálně vychýlené odhady rozptylu, které jsou garantovány jako nezáporné. Klasická analýza může vést k záporným odhadům rozptylu, které vyžadují dodatečnou akci.

**Tabulka C.9 — Softwarový výstup pro modelování smíšených vlivů**

Softwarový výstup	Poznámky
<b>Model</b>	Stručný popis vstupních dat a zvolené metody pro fitování
Lineární smíšený model, fitování pomocí REML Data: meoh Log-omezená-věrohodnost: 5,537 651	Soubor dat byl nazván „meoh“; Log-omezená věrohodnost je hodnotou fitovacího kritéria pro nejlépe vyhovující model fitování.
<b>Stálé vlivy: metanol – přibližná pozice</b>	Výraz „metanol-pozice“ pro část modelu se stálými vlivy znamená, že metanol je modelován jako jednoduchá lineární funkce pozice, přesně jako u klasické lineární regrese. Koeficienty modelu se mohou srovnávat s těmi uvedenými v tabulce C.6, i když nebudou obecně identické.
průsečík                      pozice 5,328 8                          0,004 173	
<b>Náhodné vlivy: vzorec <math>\sim 1 </math>jednotku</b>	Náhodný vliv je modelován jako jednoduchý náhodný vliv jednotky
směrodatná odchylka mezi jednotkami: 0,104 2 směrodatná odchylka uvnitř jednotek: 0,175 5	Náhodné vlivy jsou uváděny jako požadované směrodatné odchylky.
počet pozorování: 70 počet skupin: 35	Shrnutí informací o velikosti skupin pro náhodné vlivy

Směrodatná odchylka mezi jednotkami v tabulce poskytuje odhad  $s_{bb}$ , zatímco směrodatná odchylka uvnitř jednotek odpovídá  $s_w$ .



Tabulka C.10 — Výsledky pro metanol v kandidátském RM whisky

Pozice	Jednotka	Metanol mg/100ml	Reziduál mg/100ml		Pozice	Jednotka	Metanol mg/100ml	Reziduál mg/100ml
2	1 421	5,203 2	-0,138 6		40	383	5,389 4	-0,106 3
3	896	5,093 2	-0,252 7		41	1 036	5,487 3	-0,012 4
4	1 482	5,510 2	0,160 3		42	441	5,496 4	-0,007 4
5	406	5,562 5	0,208 5		43	441	5,599 2	0,091 4
6	875	5,035 2	-0,322 8		45	896	5,458 7	-0,057 2
7	1 263	5,426	0,063 9		46	729	5,611 9	0,092
8	787	5,022 3	-0,343 8		47	53	5,630 1	0,106 1
9	224	5,186 6	-0,183 6		48	956	5,646 3	0,118 3
10	687	5,548 9	0,174 7		49	813	5,377 6	-0,154 5
11	337	5,575 5	0,197 2		50	602	5,315 6	-0,220 5
13	602	5,203 9	-0,182 5		51	1 337	5,288 8	-0,251 4
14	105	5,478 1	0,087 7		52	506	5,581	0,036 8
15	1 482	5,591 3	0,196 8		53	1 129	5,129 8	-0,418 5
16	1 200	5,303 9	-0,094 6		54	1 337	5,430 2	-0,122 1
17	406	5,370 7	-0,031 9		56	1 421	5,039 3	-0,521 1
18	658	5,510 3	0,103 7		57	337	5,810 1	0,245 6
19	1 368	5,355 8	-0,054 8		58	190	5,657	0,088 5
20	1 368	5,550 8	0,136 1		59	1 309	5,407 2	-0,165 4
21	296	5,787 5	0,368 8		60	224	5,586 9	0,010 3
23	985	5,137 1	-0,289 7		61	729	5,940 4	0,359 7
24	687	5,628 6	0,197 7		62	296	5,439 4	-0,145 3
25	1 163	5,371 9	-0,063		63	1 200	5,770 3	0,181 5
26	506	5,327 1	-0,111 9		64	875	5,796 1	0,203 3
27	9	5,657	0,214		65	53	5,665	0,068 1
28	1 507	5,320 8	-0,126 3		67	105	5,640 6	0,035 6
29	9	5,864 6	0,413 5		68	151	5,494 9	-0,114 1
30	1 129	5,467 7	0,012 5		69	1 074	5,594 8	-0,018 3
31	985	5,284 7	-0,174 5		70	1 507	5,290 3	-0,326 8
32	956	5,804 1	0,340 8		71	1 074	5,628 2	0,007
34	658	5,331 3	-0,140 1		72	564	5,912 5	0,287 3
35	383	5,919 3	0,443 9		73	1 036	5,765 1	0,135 9
36	1 263	5,584 1	0,104 6		74	1 163	5,458 2	-0,175 1
37	813	5,303 9	-0,179 6		75	787	5,675 7	0,038 4
38	151	5,578 9	0,091 3		76	564	5,561 4	-0,08
39	1 309	5,308 4	-0,183 2		77	190	5,793 6	0,148 2

## Annex C (normative)

### Examples

#### C.1 Homogeneity study

A between-unit homogeneity study has been carried out in a project to prepare a certified reference material for chromium in soil. Table C.1 gives the data of the homogeneity study.

**Table C.1 — Measurement data of a between-unit homogeneity study of chromium in soil**

Bottle No.	Result No. 1	Result No. 2	Result No. 3
1	121,30	128,74	119,91
2	120,87	121,32	119,24
3	122,44	122,96	123,45
4	117,60	119,66	118,96
5	110,65	112,34	110,29
6	117,29	120,79	121,42
7	115,27	121,45	117,48
8	118,96	123,78	123,29
9	118,67	116,67	114,58
10	126,24	123,51	126,20
11	128,65	122,02	121,93
12	126,84	124,72	123,14
13	122,61	128,48	126,20
14	118,95	123,82	118,11
15	118,74	118,23	117,38
16	119,74	121,78	121,01
17	121,21	123,28	116,38
18	129,30	124,10	122,02
19	136,81	129,80	128,47
20	127,81	117,66	122,90

NOTE All results are in milligrams per kilogram

The analysis of variance shown in Table C.2 can be computed from the data in Table C.1 using, for example, spreadsheet software.

**Table C.2 — ANOVA table for between-bottle homogeneity study of chromium in soil**

Source of variation	Sum of squares	Degrees of freedom	Mean square <i>M</i>
Between bottles	1 037,1	19	54,59
Within bottles	330,5	40	8,26
Total	1 367,6	59	

The between-group variance (the between-unit variance, in a homogeneity study) is estimated using

$$s_{\text{between}}^2 = \frac{M_{\text{between}} - M_{\text{within}}}{n_0} = \frac{54,59 - 8,26}{3} \text{mg}^2\text{kg}^{-2}$$

The between-unit standard deviation is the square root of this variance

$$s_{bb} = \sqrt{15,44} = 3,93 \text{ mg kg}^{-1}$$

The repeatability standard deviation can be computed from  $M_{\text{within}}$

$$s_r = \sqrt{M_{\text{within}}} = \sqrt{8,26} = 2,87 \text{ mg kg}^{-1}$$

## C.2 Between-unit homogeneity, considering the limited repeatability of the test method

The example is taken from Reference [65], a certification of the catalytic concentration of gamma-glutamyltransferase in pig kidney tissue. Pig kidney tissue was diluted with sodium chloride and homogenized. After repeated precipitation, the product was subjected to purification by chromatography on a DEAE Trisacryl column, followed by chromatography on a hydroxyapatite column. Bovine serum albumin at a mass concentration of 60 g/l was chosen as matrix, since its addition did not modify the catalytic properties of the partially purified enzyme. The filling procedure was checked by weighing 101 ampoules taken in regular intervals. No trend over the filling time was found. Finally, the material was lyophilised.

The material was checked for contaminating enzymes, purity, water content and residual oxygen content, and was found to be suitable for a CRM.

To quantify between-unit variation, data for the assessment of homogeneity were taken from the earlier certification exercise, which included 20 ampoules measured six times each. An ANOVA over the data grouped by ampoule was performed. The within-unit standard deviation ( $s_r$ ) and between-unit standard deviation ( $s_{bb}$ ) were calculated. The results of the analysis of variance, including the mean squares and the estimates of the standard deviations at the various levels, are shown in Table C.3.

**Table C.3 — Results for catalytic concentration of gamma-glutamyltransferase in pig kidney tissue (relative values in parentheses)**

Average	67,78 IU/l
$M_{\text{between}}$	1,76 IU <sup>2</sup> /l <sup>2</sup>
$M_{\text{within}}$	1,63 IU <sup>2</sup> /l <sup>2</sup>
$s_r$	1,28 IU/l (1,88 %)
$s_{bb}$	0,147 IU/l (0,22 %)
$u'_{bb}$	0,196 IU/l (0,29 %)

The between-unit standard deviation is computed using

$$s_{bb}^2 = \max\left(\frac{M_{\text{between}} - M_{\text{within}}}{n_0}, 0\right) = \frac{1,76 - 1,63}{6} = 0,0217$$

which leads to an estimate for  $s_{bb}$  of 0,147 IU/l, equivalent to 0,22 % relative. In this instance, the within-group standard deviation has 100 degrees of freedom. The repeatability standard deviation can be computed from  $M_{\text{within}}$

$$s_r = \sqrt{M_{\text{within}}} = \sqrt{1,63} = 1,28 \text{ IU/l}$$

This was considered too large to provide good evidence of heterogeneity as low as 0,22 % relative. The producer accordingly chose to use a more conservative value for the uncertainty associated with heterogeneity, denoted  $u'_{bb}$ , based on the recommendation of Reference [66], as follows:

$$u'_{bb} = \sqrt{\frac{M_{\text{within}}}{n_0}} \cdot \sqrt{\frac{2}{v M_{\text{within}}}} = \sqrt{\frac{1,63}{6}} \sqrt{\frac{2}{100}} = 0,196 \text{ IU/l}$$

(This expression can be shown to be an estimated standard uncertainty for the within-unit standard deviation). Reference [66] recommends using the larger of  $s_{bb}$  and  $u'_{bb}$  calculated as above; the value of 0,196 IU/l was accordingly used as the uncertainty associated with the between-unit variation.

### C.3 Homogeneity assessment using a randomized block design

#### C.3.1 Background and data

A between-unit homogeneity study was carried out in a project to prepare a reference material certified for anions in river water. This example uses data for sulfate. The study was conducted as a randomized block design in which 15 units of the RM were all measured once in each of three different measurement runs. Each run was randomized separately so that the units were measured in a different order for each run. The sulfate data, ordered by RM unit number, are given in Table C.4.

**Table C.4 — Homogeneity data for sulfate (mg l<sup>-1</sup>)**

Unit	Run 1	Run 2	Run 3
15	79,737	79,036	80,196
127	80,177	78,367	80,005
199	79,950	78,555	80,040
248	78,970	79,070	79,299
309	79,164	78,865	79,567
375	79,315	77,810	80,250
439	79,720	79,142	80,424
482	80,102	78,217	80,342
571	80,457	78,569	80,022
662	80,105	78,280	80,770
689	80,550	79,333	80,530
775*	79,215	78,442	77,919
841	79,96	78,947	80,313
911	79,637	79,574	80,107
973	79,710	78,614	80,385
Run order, which was randomized, is not shown.			
* Unit 775 shows an outlying value and was discarded: see text.			

A plot of all the data in Table C.4 (Figure C.1) shows a clear run effect, together with one outlying value (identified as being associated with unit 775) in Run 3. The outlier was traced to poor chromatographic peak shape. Options for outlier removal include elimination of the individual point or elimination of the affected RM unit. In this case, to preserve a balanced design and simplify data analysis, all three replicates for the unit were discarded, leaving technically valid data for 14 units.

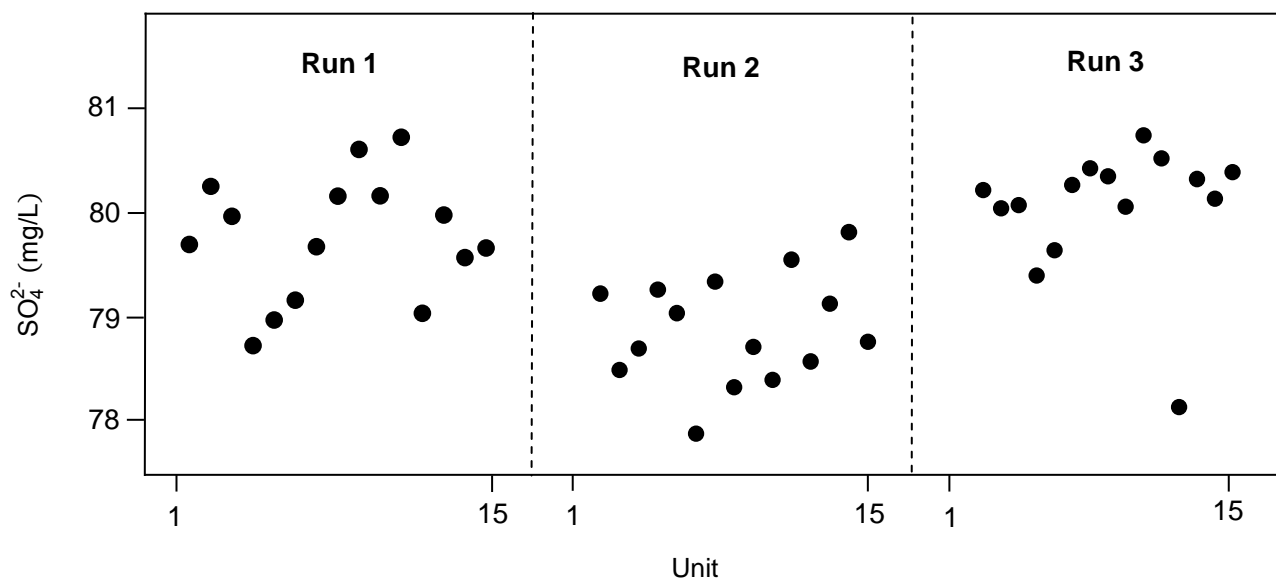
#### C.3.2 Evaluation

Two-factor (often also called "two-way") analysis of variance (ANOVA), without replication, was performed on the data in Table C.4. The resulting ANOVA table is shown as Table C.5, confirming a strongly significant Run effect.

**Table C.5 — Two-factor ANOVA table for sulfate data**

	Df <sup>a</sup>	Sum Sq <sup>b</sup>	Mean Sq <sup>c</sup>	F value <sup>d</sup>	Pr(> F) <sup>e</sup>
<b>Unit</b>	13	3,065 1	0,2358	1,3476	0,249 6
<b>Run</b>	2	15,409 3	7,704 6	44,036 4	4,48e-09
<b>Residuals</b>	26	4,549 0	0,175 0		
<sup>a</sup> Degrees of freedom <sup>b</sup> Sum of squares					

- <sup>c</sup> Mean square  
<sup>d</sup> F-statistic (mean square divided by residual mean square)  
<sup>e</sup>  $p$ -value (probability of exceeding the observed value of the F statistic under the null hypothesis of zero effect variance).



NOTE Vertical dashed lines delineate separate runs. Unit numbers 1 to 15 in this plot correspond to the 15 unit numbers in Table C.4, in unit number order.

**Figure C.1 — Observations for sulfate in a candidate reference material**

The within- and between-unit variances  $s_w^2$  and  $s_{bb}^2$  are estimated as indicated in 7.7.4:

$$s_w^2 = s_r^2 = M_w = 0,175\ 0$$

where  $M_w$  denotes the within-unit mean square, giving  $s_w = 0,418\ 3\ \text{mg l}^{-1}$  and

$$s_{bb}^2 = \max\left(\frac{M_A - M_w}{n_0}, 0\right) = \max\left(\frac{0,235\ 8 - 0,175\ 0}{3}, 0\right) = 0,020\ 3$$

where  $M_A$  denotes the between-unit mean square and  $n_0$  the number of observations per unit (identical to the number of runs in this simple randomized block design). This gives  $s_{bb} = 0,142\ 4\ \text{mg l}^{-1}$ .

NOTE For comparison, one-way analysis of variance applied to the same data grouped only by unit yields the same between-unit mean square (0,235 8) but a greatly inflated residual mean square of 0,712 8. This would lead, incorrectly, to a zero estimate of the between-unit variance  $s_{bb}$ , underestimating the uncertainty associated with homogeneity or, if an allowance for homogeneity were based on the within-unit term, leading to a substantial overestimate.

## C.4 Homogeneity assessment: simple design with correction for linear trend

### C.4.1 Background and data

A between-unit homogeneity study was carried out in a project to prepare a certified reference material for whisky congeners. 35 units of the RM were run in duplicate for homogeneity testing using a single gas chromatography run, replicates being observed in random order. Table C.10 gives the data from the homogeneity study and the residuals for a linear fit to the data (see below). In the table, unit number corresponds to fill order, while Position refers to the position in the run sequence.

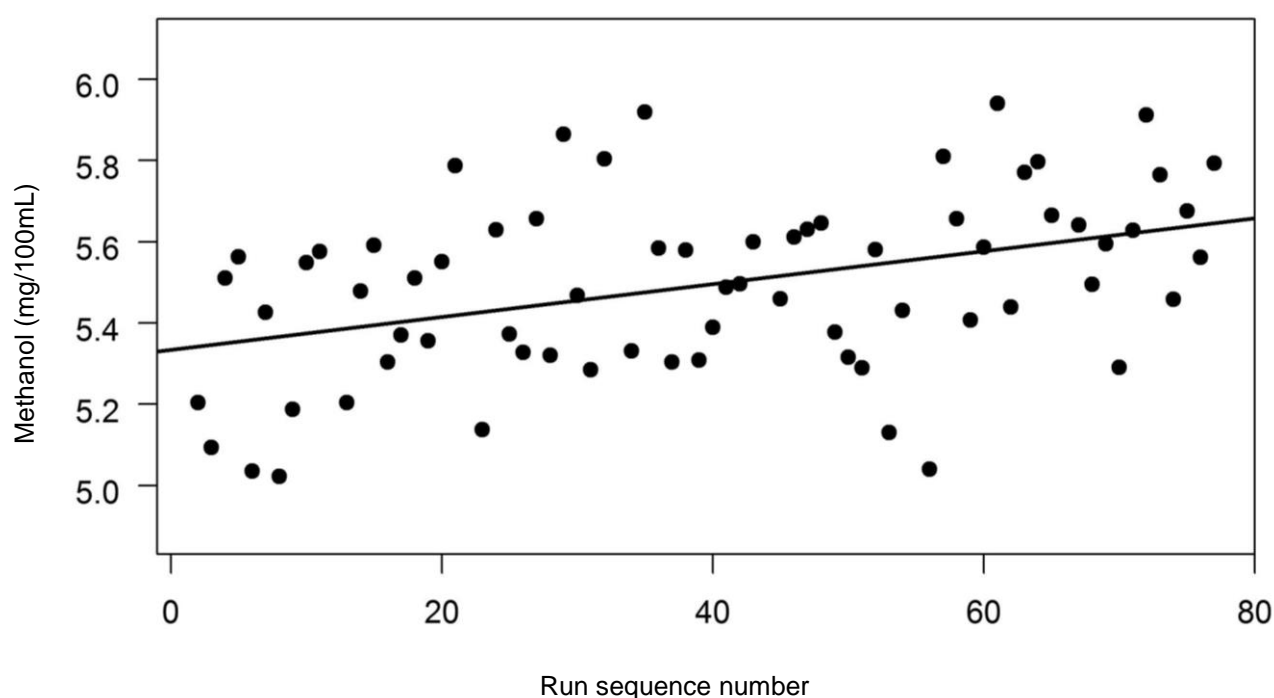
Following 7.7.1 the data were plotted in Unit and Run order. The plot in Run order is shown as Figure C.2.

A modest upward trend, attributable to instrument drift, is visible; the solid line shows the least squares straight line fitted to the data. The regression results are shown as Table C.6; they show a very strongly significant effect of Position (the slope) indicated by the small  $p$ -value (" $\text{Pr}(> |t|)$ " in the table).

**Table C.6 — Regression statistics for methanol against run sequence number**

	Estimate <sup>a</sup>	Standard error <sup>b</sup>	t value <sup>b</sup>	(Pr(>  t )) <sup>c</sup>
Intercept	5,333 7	0,049 8	107,15	< 2e-16
Position	0,004 0	0,001 1	3,69	0,000 45
<sup>a</sup> Coefficients and standard errors ("Std. error") are shown to four decimal places <sup>b</sup> $t$ -statistic (estimate divided standard error) <sup>c</sup> $p$ -value (probability of exceeding the observed absolute value of the $t$ statistic under the null hypothesis of zero gradient.				

The trend is not so strong as to demand rejection of the run and re-analysis, and visual inspection of the plot suggests that a linear fit provides a reasonable approximation. Statistical treatment is therefore used to remove the adverse effects of the trend. Two options are illustrated; correction followed by classical analysis of variance, and (for comparison of results only) the results of mixed effects modelling which incorporates the trend and the unit effect simultaneously.



NOTE The solid line is a linear fit to the data. "Run sequence number" corresponds to "Position" in Table C.10.

**Figure C.2 — Results for methanol in a candidate whisky RM plotted in run order**

#### C.4.2 Method 1: Correction for trend followed by classical ANOVA

The initial step subtracts the linear trend from the data. The results are shown as the Residual in Table C.10.

NOTE The residuals correspond to subtraction of both the intercept and position terms in the linear model. Since subsequent analysis of variance assumes additivity of unit effects and random error, it would be equally valid to subtract the position term and leave the data uncorrected for intercept.

One-way analysis of variance (ANOVA) is then applied to the residuals, with Unit as the grouping factor. The initial ANOVA table on the trend-corrected data are shown in Table C.7.

**Table C.7 — Initial ANOVA applied to whisky homogeneity residuals**

	Df <sup>a</sup>	Sum Sq <sup>b</sup>	Mean Sq <sup>c</sup>	F value <sup>d</sup>	(Pr(> F)) <sup>e</sup>
Between-unit	34	1,764 6	0,051 90	1,714 6	0,058 65
Within-unit	35	1,059 5	0,030 27	0,000 45	

<sup>a</sup> Degrees of freedom

<sup>b</sup> Sum of squares

<sup>c</sup> Mean square

<sup>d</sup> F-statistic (mean square divided by residual mean square)

<sup>e</sup> *p*-value (probability of exceeding the observed value of the F statistic under the null hypothesis of zero effect variance).

Since the residuals in Table C.10 arise from a linear model with gradient (Position) and intercept terms, however, the ANOVA table shown in Table C.7 is slightly incorrect, as the ANOVA assumes that only the grand mean is subtracted. A small correction is therefore made by

- subtracting an additional degree of freedom from the within-unit degrees of freedom in the ANOVA table to account for the gradient term and
- recalculating the residual mean square by dividing the corresponding sum of squares with the corrected number of degrees of freedom. The resulting corrected ANOVA table is shown (with the corresponding calculations) as Table C.8.

While the correction to the within-unit mean square is small, the between unit term uses the difference between mean squares and the correction may sometimes be large compared to the required difference.

If a more complex model is used for trend correction, the correction to the within-unit degrees of freedom should respect the number of parameters in the model used. Typically, if there are  $q$  parameters in the model,  $q - 1$  is subtracted from the within-group degrees of freedom.

**Table C.8 — ANOVA table corrected for linear model degrees of freedom**

	Df <sup>a</sup>	Sum Sq <sup>b</sup>	Mean Sq <sup>c</sup>	F value <sup>d</sup>	(Pr(> F)) <sup>e</sup>
Between-unit	34	1,764 6	0,051 90	1,714 6	0,058 65
Within-unit	(35–1) = 34	1,059 5	1,059 5/34 = 0,031 16		

<sup>a</sup> Degrees of freedom

<sup>b</sup> Sum of squares

<sup>c</sup> Mean square

<sup>d</sup> F-statistic (mean square divided by residual mean square)

<sup>e</sup> *p*-value (probability of exceeding the observed value of the F statistic under the null hypothesis of zero effect variance).

The determination of the within- and between-unit standard deviations  $s_{bb}$  and  $s_w$  now proceeds as in sub-clause 7.7.3:

$$s_{bb}^2 = \max\left(\frac{MS_b - MS_w}{n_0}, 0\right) = \max\left(\frac{0,051\ 90 - 0,031\ 16}{2}, 0\right) = 0,010\ 4\ \text{mg}/100\ \text{ml}$$

yielding a between-unit standard deviation  $s_{bb}$  of 0,101 8 mg/100 ml. The within-unit standard deviation  $s_w$  is just the square root of the residual mean square, which is 0,176 5 mg/100 ml.

For comparison, the estimates of  $s_{bb}$  and  $s_w$  obtained from analysis of the data without trend correction were in this instance 0,092 8 mg/100 ml and 0,201 6 mg/100 ml, respectively, slightly underestimating the between-unit standard deviation while overestimating the within-unit standard deviation.

### C.4.3 Method 2: Incorporating the run trend in the analysis

In statistical terminology, the effect of a single run trend is most commonly modelled as a so-called ‘fixed effect’, while the between-unit effect is considered to be random. The most general method of incorporating and estimating fixed and random effects together is to apply so-called “mixed effects” modelling (see, for example, Reference <sup>[26]</sup>). Since minimally biased within- and between-unit variances are of interest, the appropriate fitting criterion is restricted maximum likelihood (REML) (see 7.7.6). This requires appropriate statistical software. In the following example, the nlme package<sup>[67]</sup> in the open-source software package R <sup>[68]</sup> has been used. Since model specification and execution depend entirely on the software used, only the results are of interest. These are given, with brief comments on each part, as Table C.9.

Notice that the results agree closely with the classical calculations at C.4.2. The principal advantages are:

- Given suitable software, the calculation requires fewer operations. In this case, the results shown in Table C.9 require a single command and no further calculations. This contrasts with the separate linear model fit, ANOVA table calculation, correction of the ANOVA table and manual calculation of standard deviations required by the classical approach used in C.4.2.
- Because the model simultaneously incorporates both the random between-unit effects and the trend, the linear model better incorporates any partial confounding of unit effect position and run sequence, which may occasionally lead to over- or under-interpretation of linear regression results applied without incorporating the random error due to between-unit effects.
- Using the REML criterion provides minimally biased variance estimates that are guaranteed non-negative; the classical analysis can lead to negative variance estimates for which additional action is required.

**Table C.9 — Software output from mixed effect modelling**

Software output	Remarks
Model	A brief statement of the input data and fitting method
Linear mixed-effects model fit by REML Data: meoh Log-restricted-likelihood: 5,537 651	The data set was named “meoh”; the Log-restricted-likelihood is the value of the fitting criterion for the best fit model returned
<b>Fixed effects: Methanol approximately Position</b>	The formula “Methanol ~ Position” for the fixed-effects part of the model means that methanol is modelled as a simple linear function of position, exactly as for classical linear regression shown. The model coefficients can be compared with those in Table C.6, though they will not generally be identical
Intercept                      Position 5,328 8                              0,004 173	
<b>Random effects: Formula ~1 Unit</b>	The random effect is modelled as a simple random effect of Unit
Between-unit standard deviation:      0,104 2 Within-unit standard deviation:        0,175 5	The random effects are given as the required standard deviations
Number of Observations: 70 Number of Groups: 35	Summary information on group sizes for the random effects

The between-unit standard deviation in the table provides an estimate of  $s_{bb}$ , while the within-unit standard deviation corresponds to  $s_w$ .

**Table C.10 — Results for methanol in a candidate whisky RM**

Position	Unit	Methanol mg/100ml	Residual mg/100ml		Position	Unit	Methanol mg/100ml	Residual mg/100ml
2	1 421	5,203 2	-0,138 6		40	383	5,389 4	-0,106 3
3	896	5,093 2	-0,252 7		41	1 036	5,487 3	-0,012 4
4	1 482	5,510 2	0,160 3		42	441	5,496 4	-0,007 4



5	406	5,562 5	0,208 5		43	441	5,599 2	0,091 4
6	875	5,035 2	-0,322 8		45	896	5,458 7	-0,057 2
7	1 263	5,426	0,063 9		46	729	5,611 9	0,092
8	787	5,022 3	-0,343 8		47	53	5,630 1	0,106 1
9	224	5,186 6	-0,183 6		48	956	5,646 3	0,118 3
10	687	5,548 9	0,174 7		49	813	5,377 6	-0,154 5
11	337	5,575 5	0,197 2		50	602	5,315 6	-0,220 5
13	602	5,203 9	-0,182 5		51	1 337	5,288 8	-0,251 4
14	105	5,478 1	0,087 7		52	506	5,581	0,036 8
15	1 482	5,591 3	0,196 8		53	1 129	5,129 8	-0,418 5
16	1 200	5,303 9	-0,094 6		54	1 337	5,430 2	-0,122 1
17	406	5,370 7	-0,031 9		56	1 421	5,039 3	-0,521 1
18	658	5,510 3	0,103 7		57	337	5,810 1	0,245 6
19	1 368	5,355 8	-0,054 8		58	190	5,657	0,088 5
20	1 368	5,550 8	0,136 1		59	1 309	5,407 2	-0,165 4
21	296	5,787 5	0,368 8		60	224	5,586 9	0,010 3
23	985	5,137 1	-0,289 7		61	729	5,940 4	0,359 7
24	687	5,628 6	0,197 7		62	296	5,439 4	-0,145 3
25	1 163	5,371 9	-0,063		63	1 200	5,770 3	0,181 5
26	506	5,327 1	-0,111 9		64	875	5,796 1	0,203 3
27	9	5,657	0,214		65	53	5,665	0,068 1
28	1 507	5,320 8	-0,126 3		67	105	5,640 6	0,035 6
29	9	5,864 6	0,413 5		68	151	5,494 9	-0,114 1
30	1 129	5,467 7	0,012 5		69	1 074	5,594 8	-0,018 3
31	985	5,284 7	-0,174 5		70	1 507	5,290 3	-0,326 8
32	956	5,804 1	0,340 8		71	1 074	5,628 2	0,007
34	658	5,331 3	-0,140 1		72	564	5,912 5	0,287 3
35	383	5,919 3	0,443 9		73	1 036	5,765 1	0,135 9
36	1 263	5,584 1	0,104 6		74	1 163	5,458 2	-0,175 1
37	813	5,303 9	-0,179 6		75	787	5,675 7	0,038 4
38	151	5,578 9	0,091 3		76	564	5,561 4	-0,08
39	1 309	5,308 4	-0,183 2		77	190	5,793 6	0,148 2

“Position” corresponds to run order; QC observations have been removed. “Unit” corresponds to the RM filling sequence. “Methanol” is the value obtained for the property of interest, reported to four places. “Residual” gives the residuals for a linear fit to the data shown; residuals are truncated to four places for this table.

## Příloha D (informativní)

### Vyhodnocení nejistoty měření

Základním dokumentem, který popisuje vyhodnocování nejistoty měření je dokument ISO/IEC Návod pro vyjadřování nejistoty měření („GUM“). (513) Pokyn ISO/IEC 98-3:2008, Kapitola 8 udává podrobný postup pro slučování nejistot pocházejících z různých zdrojů nejistoty. Tento postup může být shrnut následovně:

- a) Vyjádřete, matematicky, vztah mezi hodnotou vlastnosti, která má být certifikována a všemi vstupními veličinami, na nichž závisí hodnota vlastnosti.
- b) Stanovte hodnoty všech vstupních veličin, založené buď na statistické analýze řady dat, nebo získané jinými prostředky.
- c) Vyhodnoťte standardní nejistotu všech vstupních veličin pomocí statistické analýzy dat, nebo jinými způsoby (jako třeba odkazem  $n$  a certifikáty, profesionálním posudkem).
- d) Vyhodnoťte kovariance mezi jakýmkoli vstupními veličinami.
- e) Vypočtete hodnotu vlastnosti ( $x$ ), tj. hodnotu vlastnosti, která má být certifikována.
- f) Stanovte kombinovanou standardní nejistotu přidruženou hodnotě vlastnosti ze standardních nejistot a kovariancí přiřazených vstupním veličinám.
- g) Stanovte interval pokrytí s udanou hladinou spolehlivosti (často 95%). V případě symetrických rozdělení, interval pokrytí může být udán ve formě rozšířené nejistoty, získané vynásobením standardní nejistoty přidružené k hodnotě vlastnosti koeficientem  $k$ .
- h) Hodnota vlastnosti by měla být udávána spolu s jejím intervalem pokrytí, obvykle s udáním spodní a horní hranice intervalu. Za účel šíření nejistoty by měla být rovněž uvedena standardní nejistota přidružená k hodnotě vlastnosti. V případech, kdy je rozdělení vlastnosti symetrické, tento interval pokrytí může být vyjádřen jako rozšířená nejistota  $U$  a koeficient rozšíření  $k$ .

V mnoha případech, může být následován shora popsán postup. Existují však situace, kdy by měly být zvoleny jiné přístupy diskutované v GUM [51]. Tyto situace zahrnují

- případy, kdy neexistuje vhodný (těsný) matematický výraz pro model popisující vztah mezi hodnotou vlastnosti a vstupními veličinami a
- případy, kdy lineární aproximace, získaná použitím vztahu pro kombinování a šíření nejistot, je jasně neplatná.

Pro nelineární modely měření by měla být potvrzena platnost zákona o šíření nejistot. V případě pochyby by měly být místo toho použity metody Monte Carlo, popsané v Dodatku č.1 GUM [5].

Když je certifikován soubor vlastností, které sdílejí společný vliv veličin (např. certifikované hodnoty série kalibračních standardů o různé koncentraci odvozené ze stejného zásobního roztoku), korelace mezi různými hodnotami může být dostatečně velká, takže je nutné poskytnout konečnému uživateli informaci o kovarianci. V těchto případech, vícevariantní přístup k vyhodnocování nejistoty může být použit, aby se získala kovarianční matice přidružená k souboru hodnot vlastnosti. Dodatek č.2 GUM poskytuje podrobnosti tohoto přístupu.

Podrobnosti jak vyhodnocovat jednotlivé složky nejistoty jsou uvedeny v GUM [51] a jeho dodatcích. V mnoha případech lze definovat sloučené příspěvky nejistoty, aby se s výhodou využily všechny existující údaje, jako jsou výsledky validačních studií, viz Kapitola 7, Reference [52].

## Annex D (informative)

### Measurement uncertainty evaluation

The principal document that describes the evaluation of measurement uncertainty is the ISO/IEC Guide to the expression of uncertainty in measurement (“the GUM”).<sup>[51]</sup> ISO/IEC Guide 98-3:2008, Clause 8 provides a detailed procedure for combining uncertainties arising from different sources of uncertainty. This procedure can be summarized as follows.

- a) Express, mathematically, the relationship between the property value to be certified and all input quantities on which the property value depends. The relationship should include all quantities that might contribute significantly to the uncertainty of the property value, and is called the measurement model.
- b) Determine the values for all input quantities, based either on statistical analysis of a series of data or by other means.
- c) Evaluate the standard uncertainty for all input quantities using a statistical analysis of data, or by other means (such as reference to certificates or professional judgement).
- d) Evaluate the covariances between any input quantities.
- e) Calculate the property value ( $x$ ), i.e. the value of the property to be certified.
- f) Determine the combined standard uncertainty associated with the property value from the standard uncertainties and covariances associated with the input quantities.
- g) Determine a coverage interval with a specified level of confidence (often 95 %). In the case of symmetric distributions, the coverage interval can be stated in the form of an expanded uncertainty, obtained by multiplying the standard uncertainty associated with the property value by a coverage factor  $k$ .
- h) The property value should be reported together with its coverage interval, usually by stating the lower and upper bounds of the interval. For the purpose of propagating uncertainty, the standard uncertainty associated with the property value should be given as well. In cases where the distribution of the property is symmetric, this coverage interval can be stated as an expanded uncertainty  $U$  and the coverage factor  $k$ .

In many cases, the approach as described above can be followed. There are however some situations, where other approaches should be chosen, as discussed in the GUM.<sup>[51]</sup> Such situations include

- cases where there is no closed mathematical form for the model describing the relationship between the property value and the input quantities, and
- cases where the linear approximation, as obtained by applying the formula for combining and propagating uncertainties, is clearly invalid.

For nonlinear measurement models, the validity of applying the law of propagation of uncertainty should be confirmed. In case of doubt, the Monte Carlo methods as described in the GUM Supplement 1<sup>[5]</sup> should be used instead.

When certifying a set of properties that share common influence quantities (for example, the certified values on a set of calibration standards at different concentrations derived from the same stock solution), correlation between the different values can be sufficiently large that it is necessary to provide end users with covariance information. In these cases, a multivariate approach to the uncertainty evaluation can be used in order to obtain the covariance matrix associated with the set of property values. GUM Supplement 2<sup>[6]</sup> provides details of this approach.

The details of how to evaluate single components of uncertainty are covered by the GUM<sup>[51]</sup> and its supplements. In many cases, aggregated uncertainty components can be defined in order to take advantage of existing data, such as the results of validation studies, as described in Chapter 7 of Reference<sup>[52]</sup>.

## Bibliografie

- [1] ISO Guide 33:2015, *Reference materials — Good practice in using reference materials*
- [2] ISO Guide 34:2009, *General requirements for the competence of reference material producers*
- [3] ISO Guide 31:2015, *Reference materials — Contents of certificates, labels and accompanying documentation*
- [4] ISO/IEC Guide 99:2007, *International vocabulary of metrology — Basic and general concepts and associated terms (VIM)*
- [5] ISO/IEC Guide 98-3:2008/Suppl.1:2008, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)/Supplement 1: Propagation of distributions using a Monte Carlo method*
- [6] ISO/IEC Guide 98-3:2008/Suppl.2:2011, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)/Supplement 2: Extension to any number of output quantities*
- [7] Doc. BCR/48/93, "Guidelines for the production and certification of BCR reference materials. European Commission, Brussels, 1994
- [8] ISO 2859 (all parts), *Sampling procedures for inspection by attributes*
- [9] ISO/REMCO POSITION PAPER. The need for assessment of commutability of reference materials", International Organization for Standardization (ISO), Geneva, 2014, [http://isotc.iso.org/livelink/livelink/fetch/8854933/8854951/8854960/279217/Commutability\\_document\\_fin\\_al.pdf?nodeid=16787892&vernum=-2](http://isotc.iso.org/livelink/livelink/fetch/8854933/8854951/8854960/279217/Commutability_document_fin_al.pdf?nodeid=16787892&vernum=-2)
- [10] ISO 13528:2015, *Statistical methods for use in proficiency testing by interlaboratory comparison*
- [11] BOX G.E.P., HUNTER J.S., HUNTER W.G. *Statistics for Experimenters: Design, Innovation, and Discovery*", Second edition, J. Wiley and Sons
- [12] ELLISON S.L.R., GRIMES N., FARDUS F., ELLISON M.P.L. *Use of a replicated Latin square design in a homogeneity test for high purity organic melting point standards*", 2012, Accreditation and Quality Assurance, Vol. 17, s. 283-290
- [13] VAN DER VEEN A.M.H., LINSINGER T.P.J., LAMBERTY A., PAUWELS J. *Uncertainty calculations in the certification of reference materials 4. Characterisation and certification*", 2001, Accreditation and Quality Assurance, Vol. 6, s. 290-294
- [14] ISO 5725-3:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method*
- [15] SEARLE S.R., & CASELLA G. MCCULLOCH, C.E. *Variance components*. J. Wiley and Sons, 2006
- [16] LINSINGER T.P.J., PAUWELS J., VAN DER VEEN A.M.H., Schimmel H., Lamberty A. *Homogeneity and Stability of Reference Materials*", 2001, Accreditation and Quality Assurance, Vol. 6, s. 20-25
- [17] INGAMELLS C.O., & SWITZER P. *A proposed sampling constant for use in geochemical analysis*. Talanta. 1973, 20 s. 547-568
- [18] VAN DER VEEN A.M.H., LINSINGER T.P.J., LAMBERTY A., PAUWELS J. *Uncertainty calculations in the certification of reference materials. 2. Stability study*", 2001, Accreditation and Quality Assurance, Vol. 6, s. 257-263
- [19] LAMBERTY A., SCHIMMEL H., PAUWELS J. *The study of the stability of reference materials by isochronous measurements*" 1997, Fresenius Journal of Analytical Chemistry, Vol. 360, s. 359-361
- [20] ICH Q1D, *"Bracketing and matrixing designs for stability testing of new drug substances and products"*, International Conference on Harmonisation, 2002
- [21] TYDEMAN M.S., & KIRKWOOD T.B.L. Design and analysis of accelerated degradation tests for the stability of biological standards II. A flexible computer program for data analysis. J. Biol. Stand. 1984, 12 s. 207-214
- [22] BREMSER W., BECKER R., KIPPHARDT H., LEHNIK-HABRINK P., PANNE U., TÖPFER A. *Stability testing in an integrated scheme*", 2006, Accreditation and Quality Assurance, Vol. 11, s. 489-495

- [23] BECKER R., BUGÉ H.-G., BREMSER W., Nehls I. *Mineral oil content in sediments and soils: comparability, traceability and a certified reference material for quality assurance*, 2006, Analytical and BioAnalytical Chemistry, Vol. 385, s. 645-651
- [24] KOCH M., BREMSER W., KÖPPEN R., TÖPFER A., NEHLS I. *Development of two Certified Reference Materials for Acrylamide Determination in Foods*, 2009, Journal of Agriculture and Food Chemistry, Vol. 57, s. 8202-8207
- [25] WEST B.T., WELCH K.B., GAŁECKI A.T. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman and Hall/CRC, 2006
- [26] PINHEIRO J., & BATES D. *Mixed-Effects Models in S and S-PLUS*. Springer, 2000
- [27] LINSINGER T.P.J., & EMONS H.E. *Characterization of reference materials: proposal for a simplification of the options listed in ISO Guide 34*, 2013, Accreditation and Quality Assurance, Vol. 18, s. 149-152
- [28] EURACHEM/CITAC Guide. *Traceability in Chemical Measurement*. EURACHEM, 2003
- [29] ISO 6865:2000, *Animal feeding stuffs — Determination of crude fibre content — Method with intermediate filtration*
- [30] ISO 6142-1:2015, *Gas analysis — Preparation of calibration gas mixtures — Part 1: Gravimetric method for Class I mixtures*
- [31] MILTON M.J.T., VARGHA G.M., BROWN A.S. Gravimetric methods for the preparation of standard gas mixtures. *Metrologia*. 2011, 48 s. R1-R10
- [32] VAN DER VEEN A.M.H. & HAFNER K. Atomic weights in gas analysis. *Metrologia*. 2014, 51 s. 80-86
- [33] RUHKIN A.L., & VANGEL M.G. Estimation of a Common Mean and Weighted Means Statistics. *J. Am. Stat. Assoc.* 1998, 93 s. 303-308
- [34] CCQM Guidance note, *“Estimation of a consensus KCRV and associated Degrees of Equivalence”*. Consultative Committee for Amount of Substance (CCQM), Bureau International des Poids et Mesures (BIPM), Sèvres, Paris, 2013
- [35] COX M. The evaluation of key comparison data. *Metrologia*. 2002, 39 str. 589–595
- [36] VAN DER VEEN A.M.H. *Determination of the certified value of a reference material appreciating the uncertainty statements obtained in the collaborative study*. [book auth.] P, COX, M.G., FILIPE, E. PAVESE, F., RICHTER, D. CIARLINI. *Advanced Mathematical and Computational Tools in Metrology*, Singapore, World Scientific, 2000, s. 326-340
- [37] PAUWELS J., LAMBERTY A., SCHIMMEL H. The determination of the uncertainty of reference materials certified by laboratory intercomparison. *Accredit. Qual. Assur.* 1998, 3 s. 180-184
- [38] PAULE R C, & MANDEL J. *Consensus Values and Weighting Factors*, 1982, *J Res Natl Inst Stand Technol*, Vol. 87, s. 377-385
- [39] SAITO T., IHARA T., KOIKE M., KINUGASA S., FUJIMINE Y., NOSE K., HIRAI T. A new traceability scheme for the development of international system-traceable persistent organic pollutant reference materials by quantitative nuclear magnetic resonance. 2009, *Accreditation and Quality Assurance*, Vol. 14, s. 79-86
- [40] SAITO T. et al. Efficient production of reference materials of hazardous organics using smart calibration by nuclear magnetic resonance. *Accredit. Qual. Assur.* 2011, 16 s. 421-428
- [41] WESTWOOD S. et al. Mass balance method for the SI value assignment of the purity of organic compounds. *Anal. Chem.* 2013, 85 s. 3118-3126
- [42] ICH Q3a, *“Harmonised tripartite guideline Q3A(R2) Impurities in new drug substances”*, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) ICH, [online] 2006, [Citováno: 11 března 2013] <http://www.ich.org>
- [43] EUR 22504, *“Certification of a reference material of purified genomic DNA from Campylobacter jejuni (NCTC 11351) IRMM-448”*, 2007
- [44] European Pharmacopoeia. Council of Europe, Strasbourg, Eighth Edition, 2014

- [45] HAYNES W.M. CRC Handbook of Chemistry and Physics. CRC Press, Boca Raton, Ninety fifth Edition, 2014 [ed.]
- [46] MCNAUGHT A.D., & WILKINSON A. [ed.]. IUPAC. Compendium of Chemical Terminology, International Union of Pure and Applied Chemistry, 2014. XML online corrected version created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins, ISBN 0-9678550-9-8
- [47] 2002/657/EC, Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (Text with EEA relevance) (notified under document number C(2002) 3044)
- [48] ISO 11843-1, *Capability of detection — Part 1: Terms and definitions*
- [49] JOSEPHS R., KOEBER R., BERNREUTHER A., SCHIMMEL H., ULBERTH F. EUR Report 21202 EN: The certification of the mass fraction of aflatoxin M1 in whole milk powders Certified Reference Material ERM®-BD282 (zero level) ERM®-BD283 (low level) ERM®-BD284 (high level). Joint Research Centre, European Commission, 2005
- [50] HUNTER R.S. *Accuracy, precision and stability of new photoelectric colour-difference meter*. J. Opt. Soc. Am. 1948, 38 s. 1094A
- [51] ISO/IEC Guide 98-3:2008, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)*
- [52] ELLISON S.L.R., & WILLIAMS A. eds. *“Eurachem/CITAC guide: Quantifying Uncertainty in Analytical measurement”*, 2012
- [53] ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories*
- [54] LEVENSON M.S., BANKS D.L., EBERHARDT K.R., GILL L.M., GUTHRIE W.F., LIU H.K. et al. An approach to combining results from multiple methods motivated by the ISO GUM. J. Res. Natl. Inst. Stand. Technol. 2000, 105 s. 571-579
- [55] TOMAN B., & POSSOLO A. Laboratory effects models for interlaboratory comparisons. Accred. Qual. Assur. 2009, 14 s. 553-563
- [56] VAN DER VEEN A.M.H., & PAUWELS J. Uncertainty calculations in the certification of reference materials. 1. Principles of analysis of variance. Accreditation and Quality Assurance, 2000, 5, s. 464-469
- [57] SOKAL R.R., & ROHLF F.J. Biometry. Freeman, New York, Third Edition, 1995
- [58] JOHN P.W.M. Statistical design and analysis of experiments. Macmillan, New York, 1971
- [59] SNEDECOR G.W., & COCHRAN W.G. Statistical Methods. Iowa State University Press, Eighth Edition, 1989
- [60] SEARLE S.R. Linear Models. John Wiley, New York, 1971
- [61] SEARLE S.R., CASELLA G., MCCULLOCH C.E. Variance Components. John Wiley & Sons, Inc, Hoboken, NJ, 2006
- [62] GRIEPINK B., MAIER E., MUNTAU H., WELLS D.E. *The certification of the contents of six chlorobiphenyls (Nos 28, 52, 101, 108, 153 and 180) in dried sludge (CRM 392)*, European Commission Joint Research Centre, Luxembourg, 1990. Certification report: EUR 12823
- [63] VAN DER VEEN A.M.H., LINSINGER T.P.J., PAUWELS J. *Uncertainty calculations in the certification of reference materials 2. Homogeneity study*”, 2001, Accreditation and Quality Assurance, Vol. 6, s. 26-30
- [64] MONTGOMERY D.C. Design and Analysis of Experiments. John Wiley & Sons, Hoboken, NJ, 2012
- [65] SCHIELE F., SIEST G., MOSS D.W., COLINET E. *The certification of the catalytic concentration of gamma-glutamyltransferase in a reconstituted lyophilized material (CRM 319)*”, European Commission, Joint Research Centre, 1986. Certification report. EUR 10628
- [66] LINSINGER T.P.J., PAUWELS J., LAMBERTY A., SCHIMMEL H., VAN DER VEEN A.M.H., SIEKMANN L. *Estimating the uncertainty of stability for matrix CRMs*”, 2001, Fresenius Journal of Analytical Chemistry, Vol. 370, s. 183-188
- [67] PINHEIRO J., BATES D., DEBROY S., SARKAR D., and the R DEVELOPMENT CORE TEAM *Linear and Nonlinear Mixed Effects Models. R package version 3.1-103*. 2012

- [68] R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012
- [69] RICE J.A. Mathematical statistics and data analysis. Duxbury Press, Belmont, Second Edition, 1995
- [70] ISO Guide 30:2015, *Reference materials — Selected terms and definitions*
- [71] IUPAC. *Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")*, Blackwell Scientific Publications, Oxford, 1997
- [72] MILTON M.J.T. The mole, amount of substance and primary methods . *Metrologia*. 2013, 50, s. 158-163
- [73] ISO 15189, *Medical laboratories — Requirements for quality and competence*