



CNAS-GL017

**标准物质/标准样品定值的
一般原则和统计方法**

**Reference materials - General and statistical principles
for certification
(ISO Guide35:2006)**

中国合格评定国家认可委员会

目 次

前 言	3
引 言 (ISO)	4
1 范围	5
2 规范性引用文件	5
3 术语和定义	6
4 符号	7
5 定值项目设计	8
5.1 总则	8
5.2 项目定义	8
5.3 运输	9
5.4 原材料收集	9
5.5 可行性研究	9
5.6 要求的使用寿命和有效期	10
5.7 样品制备	10
5.7.1 概述	10
5.7.2 人工合成材料	10
5.7.3 混合	10
5.7.4 加料	11
5.7.5 均匀化与分装	11
5.8 均匀性研究	11
5.9 稳定性研究	12
5.10 测量方法选择	13
5.11 证书	14
5.12 项目设计小结	14
6 测量不确定度评估	14
6.1 CRM/RM 特性值不确定度评估基础	14
6.2 批测定的基本模式	15
6.3 不确定度来源	17
6.4 分布函数问题	17
6.5 比率的运用	17
6.6 包含因子选择	18
6.7 重新定值	18
7 均匀性研究	18
7.1 概述	18
7.2 材料	19
7.3 均匀性的概念	19
7.4 实际应用	20
7.5 测量	20
7.6 统计上有效的抽样方案和趋势分析	20
7.7 均匀性研究评估	21
7.8 瓶间均匀性研究	21
7.9 测量方法重复性欠佳的情况	23
7.10 瓶内均匀性	23
8 稳定性研究	24
8.1 稳定性类型	24
8.2 实验设计	25
8.3 结果评估	26
8.3.1 趋势分析	26

8.3.2 无趋势时不确定度的评估	28
8.4 稳定性监测	30
8.4.1 实验	30
8.4.2 不确定度评估	31
8.5 确定与长期稳定性相关的有效期	32
9 特性值测定	32
9.1 总则	32
9.2 建立和证明溯源性	32
9.3 实用方法	35
9.4 测量方案设计	35
9.4.1 一个实验室用一种或多种（标准）方法测量	35
9.4.2 多个实验室协作研究	36
9.5 有关特性的一些问题	39
9.5.1 物理特性 RM	39
9.5.2 化学特性 RM	40
9.5.3 约定特性的测定	42
10 数据及不确定度评估	42
10.1 模型	42
10.2 数据格式	43
10.3 分布	44
10.4 数据审核	45
10.5 数据评估	45
10.5.1 方式 A: 单个实验室单一方法	45
10.5.2 方式 C: 方法和/或实验室网络	45
10.5.3 方式 B: 单个实验室的多种方法	46
10.5.4 方式 D: 方法定义的参数	46
10.5.5 异常数据的处理	46
10.6 不确定度评估	47
10.6.1 方式 A: 单个实验室的单一方法	47
10.6.2 方式 C: 方法和/或实验室网络	47
10.6.3 方式 B: 单个实验室的多种方法	47
10.6.4 方式 D: 方法定义的参数	47
10.7 基于不确定度的评估	47
10.7.1 基础	47
10.7.2 χ^2 拟合	48
10.8 一些特定问题	49
10.8.1 应用方差分析法进行数据评估	49
10.8.2 应用协作研究进行均匀性确认	49
10.8.3 特性值的合成标准不确定度	50
11 证书	50
附录 A（资料性附录）统计方法	52
附录 B（资料性附录）实例	57
参考文献	66

前 言

本文件等同采用ISO Guide35:2006 Reference materials — General and statistical principles for certification (GB/T 15000.3-2008)。作为CNAS对标准物质/标准样品生产者能力认可的指南，取代CNAS-CL04:2007《标准物质/标准样品生产者能力认可准则》第二部分“标准物质/标准样品的定值——通用原则和统计原理”。

本文件虽然为指南性文件，不作为对RMP认可的要求，但CNAS鼓励各RMP按照本文件中所述原理和方法，对RM进行定值，以确保RM质量。

引 言 (ISO)

标准物质/标准样品 (RM) 的生产、测定和定值是改善和维持世界范围测量一致性体系的关键活动。正如ISO指南32和33中所指出的,有证标准物质/标准样品 (CRM) 主要用于校准、质量控制和方法验证目的,也用于给其它材料赋值,这些材料反过来又可成为CRM;而且,CRM还用于建立或维持诸如辛烷值、硬度和pH等约定标尺的溯源性;此外还有很重要的一点,所选的纯物质可用于维持国际温度标尺。

现有3个ISO指南帮助CRM生产者建立一个生产和测定RM以及确保生产的CRM质量符合最终使用者要求的设施。ISO指南34 (CNAS-CL04) 概述了CRM生产者证明其能力应满足的要求,该部分也为如何符合这些要求提供了帮助。本指南从最通用的角度提供了对候选CRM进行均匀性检验、稳定性检验和测定的模型。ISO指南31 (CNAS-GL010) 描述了CRM证书的格式和内容。

本指南在某些方面可看作是测量不确定度表述指南 (GUM) 在CRM生产这一特殊领域内的应用。只要可能,本指南参考GUM,因为后者详细阐述了如何评估测量值的测量不确定度。本指南在某种意义上是对GUM的补充,即提供了有关CRM特性值不确定度中包含由(剩余)批不均匀性、不稳定性引起的不确定度和对这些不确定度贡献测定的附加指南。

尽管制定本指南是为更好地生产和测定RM提供支持,但如果在使用本指南时,不仔细考虑本指南中的某些特定部分是否适用于特定的CRM,那么,仍有可能使其特性值(和不确定度)建立在一个错误或有问题的基础之上。文件的使用者应注意该文件不能替代“批判性思考、理性诚实和专业技能”(GUM:1993, 3.4.8)。CRM“产品”的质量对这些方面的依赖不亚于采用适当的程序和方法。

在一个典型的定值项目中,为了正确地进行定值和解释实验数据,既要具备材料及其特性以及均匀性检验、稳定性检验和材料测定中所用测量方法的全面知识,也要具备统计方法方面的全面知识。这些必需的技能组合使得RM生产和定值变得非常复杂,其中最大的挑战是将这些技能组合起来使项目计划平稳实施。

本指南的大部分内容可用于RM生产。诸如特性值的溯源性、测量不确定度全面评估的必要性等要求特别适用于那些作为校准器、或核查方法性能的工具、或给另一个材料赋值的绝大部分RM。

药典标准和药物由药典权威机构按照本指南的一般原则建立和发布,现已有这类RM生产的专门指南。但应注意,药典权威机构采用不同的方式通过分析证书和有效日期向用户提供信息,并且不说明赋值的不确定度,因为在有关这些RM用法的摘要中不允许对其进行规定。

标准物质/标准样品定值的一般原则和统计方法

1 范围

本指南给出的统计原理旨在帮助理解和制定为标准物质/标准样品特性赋值的有效方法，包括评估有关不确定度和建立计量溯源性的方法。按照本指南描述的所有步骤生产的标准物质/标准样品（RM）通常附有证书、并被称为有证标准物质/标准样品（CRM）。本指南将有助于充分发挥CRM的潜力，以确保按国家或国际标度使测量结果具有可比性、准确性和一致性。

为了能在时空上可比，测量需要溯源到适当和规定的测量标准。在化学、生物学和物理学尤其是涉及材料和/或样品的学科中，CRM在建立测量结果溯源性方面具有重要作用。实验室应用CRM作为易于获得的测量标准建立其测量结果与国际标准的溯源性。在CRM生产过程中，其特性值可以溯源到SI单位或其它国际协议单位。本指南阐明如何制定出能很好地确定特性值的方法，使其可溯源到适当和规定的测量标准。本指南适用于从混合气体到生物样品一个很广泛的材料（基质）范围，也适用于从化学成分到物理和免疫检验特性一个非常宽的特性范围。

本指南所描述的方法不一定适用于RM生产和特性值（包括有关不确定度）确定的各个方面。可以认为本指南所给出的方法是大部分RM生产和赋值的主要方法，对一些特殊情况可能需要进行适当的修正。本指南通过举例说明了其所规定的统计方法的概要，并假定数据为正态分布。当数据确定为非正态分布时，最好采用其它统计方法以获得有效的特性值及相应的不确定度。本指南概述了生产CRM方案的设计。

2 规范性引用文件

下列文件中的条款通过本指南的引用而成为本指南的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本指南，然而，鼓励根据本指南达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 3358.1—1993 统计学术语 第一部分 一般统计术语

ISO指南30 与标准物质/标准样品有关的术语和定义

GB/T 15000.4—2003 标准物质/标准样品工作导则（4）标准物质/标准样品证书和标签的内容

GB/T 6379.1—2004 测量方法与结果的准确度（正确度与精密度）第1部分：总则与定义

GB/T 6379.2—2004 测量方法与结果的准确度（正确度与精密度）第2部分：确定标准测量方法重复性与再现性的基本方法

ISO 5725-3—2003 测量方法和测量结果的准确度（正确度与精密度）第3部分：

标准测试方法精密度的中间度量

GB/T 6379.4-2006 测量方法与结果的准确度（正确度与精密度）第4部分：确定标准测量方法正确度的基本方法

ISO 5725-5-2002 测量方法和结果的准确度（正确性和精密度）第5部分：标准测量方法的精确度测定的替代方法

ISO 5725-6-2002 测量方法和结果的准确度（正确度与精密度）第6部分：准确度值的实际应用

测量不确定度表述指南，BIPM，IEC，IFCC，ISO，IUPAC，IUPAP，OIML，1993¹⁾

计量学中基础和通用术语的国际词汇，BIPM，IEC，IFCC，ISO，IUPAC，IUPAP，OIML，1993

注：此后将“测量不确定度表述指南”用“GUM”表示，而将“计量学中基础和通用术语的国际词汇”用“VIM”表示。

3 术语和定义

GB/T 3358.1-1993、ISO指南30和VIM中确定的术语和定义以及下面将要给出的术语和定义适用于本指南。所用的符号将在第4章给出。

3.1 标准物质/标准样品 (RM) reference material

一种或多种规定特性足够均匀和稳定的材料，已被确定其符合测量过程的预期用途的。

注 1：RM是一个通用术语。

注 2：特性可以是定量或定性的（例如：物质或物种的属性）。

注 3：用途可包括测量系统校准、测量程序评估、给其它材料赋值和质量控制。

注 4：在给定测量中一种RM只能有一种用途。

3.2 有证标准物质/标准样品 (CRM) certified reference material

采用计量学上有效程序测定了一个或多个规定特性值的标准物质/标准样品，并附有证书提供规定特性值及其不确定度和计量溯源性的陈述。

注 1：值的概念包括定性特征如属性或顺序，这种特征的不确定度可用概率来表示。

注 2：标准物质/标准样品生产和定值所采用的计量学上有效程序已在GB/T 15000.7和本指南等中给出。

注 3：GB/T 15000.4给出证书内容的指南。

3.3 (标准物质/标准样品的) 特性值 property value (of a reference material)

赋予(有证)标准物质/标准样品的物理、化学或生物特性量的值。

3.4 (标准物质/标准样品的) 测定 characterization (of a reference material)

确定标准物质/标准样品特性值的过程，是定值过程的一部分。

注 1：测定过程提供被定量特性的值。

1) 该版本在 1995 年进行过修正和重印。

注 2: 批定值时, 测定的是批的特性值。

3.5 瓶间均匀性 between-bottle homogeneity

标准物质/标准样品的特性在瓶与瓶之间的变差。

注: 术语“瓶间均匀性”适用于其它类型的包装(例如小瓶)和其它物理形状及试片。

3.6 瓶内均匀性 within-bottle homogeneity

标准物质/标准样品的特性在瓶内的变差。

3.7 混合 blending

两种或多种基体材料混合以得到一种具有特定性质的材料。

3.8 基体材料 matrix material

从自然界、工业生产或其它地方采样得到的材料。

例如: 土壤、饮用水、空气。

3.9 加料 spiking

向基体材料中加入已知量的化合物或元素。

3.10 短期稳定性 short-term stability

在规定运输条件下标准物质/标准样品特性在运输过程中的稳定性。

3.11 长期稳定性 long-term stability

在CRM生产者规定贮存条件下标准物质/标准样品特性的稳定性。

3.12 (标准物质/标准样品的) 使用寿命 life time (of a reference material)

标准物质/标准样品可被使用的时间间隔。

3.13 (RM/CRM的) 有效期 shelf life (of an RM/CRM)

生产者保证CRM稳定的时间间隔。

(RM/CRM的) 有效期与GB/T 15000.4—2003中证书的有效期相同。

4 符号

A_i	方差分析(ANOVA)的偏移项
a	方差分析(ANOVA)的组数
B_i	方差分析(ANOVA)的偏移项
b	方差分析(ANOVA)的子组数
ϵ	方差分析(ANOVA)的误差项 ²⁾
k	包含因子
MS	方差分析(ANOVA)的均方
n	观测数
n_0	(有效)(子)组单元数
p	合作研究中实验室数
s_{bb}	瓶间(不)均匀性标准偏差
s_{lor}	因缺乏重复性引起的标准偏差
s_{lts}	长期(不)稳定性标准偏差
s_r	重复性标准偏差
s_{stab}	因(不)稳定性引起的标准偏差
s_{sts}	短期(不)稳定性标准偏差
s_{wb}	瓶内标准偏差

²⁾ 本指南在严格统计意义上使用误差术语, 即指观测值与其数学期望之间的差。

SS	方差分析 (ANOVA) 的平方和
u_{bb}	因瓶间 (不) 均匀性引起的标准不确定度
u_{char}	因测定引起的标准不确定度
u_{CRM}	特性值的标准不确定度
u_{lts}	因长期 (不) 稳定性引起的标准不确定度
u_{sts}	因短期 (不) 稳定性引起的标准不确定度
U_{CRM}	特性值的扩展不确定度
x_{char}	测定得到的特性值
x_{CRM}	CRM的特性值
δx_{bb}	表示瓶间 (不) 均匀性的误差项
δx_{lts}	表示长期 (不) 稳定性的误差项
δx_{sts}	表示短期 (不) 稳定性的误差项
$x_{i,j}$	方差分析 (ANOVA) 试验中单次测量的结果
μ	总体平均值 (期望)

注 1: 在某些条款中, 符号用来举例说明用于解决值项目中统计问题的典型方法。这些符号在文中解释。

注 2: 符号MS和SS采自文献, 与ISO有关符号使用的规则不相符, 但由于其概念很清晰, 认为科学文献中的习用符号适用于本指南。。

5 定值项目设计

5.1 总则

在开展CRM生产的实际活动之前需要进行大量的项目策划工作, 其中主要涉及所需材料的数量以及均匀性、稳定性和测定研究的设计, 也包括为这些研究选择适当的测量方法。样品的生产数量是策划过程中一个非常重要的变量。样品的数量和所需原材料的量取决于以上所有这些因素。在以下均匀性研究 (见第7章)、稳定性研究 (见第8章) 和测定 (见第9、10章) 的各章中, 为如何策划和实施项目策划提供了指导。可行性研究也是项目策划的一部分。

5.2 项目定义

项目策划从将要生产什么CRM开始, 一个典型的例子如下所述:

“制备一种包含系列痕量元素、适于环境分析化学相应含量水平、标准值的不确定度小于或等于x%的土壤CRM”

这个定义很好地概括了项目的内容。与环境化学相关的项目可能各不相同, 但是该定义充分限定了材料的范围, 同样, “土壤”也限定了基体的选择范围。总之, 详细说明将要生产什么是非常重要的。在项目设计阶段可以更详细地阐明定义。最终指定的目标不确定度可确保该材料将符合预期用途。例如, 校准标准值的不确定度应远小于确认痕量环境分析方法所用材料的值的不确定度。

合理选择特性值溯源性溯源到的“规定测量标准”是设计的主要问题, 这主要取决于可以得到什么测量标准、什么是实验室常规测量所必需的特定CRM、以及什么是技术上可行的。由于CRM主要使后续测量具有可溯源性, 因此, 选择合适的测量标准

对于所生产的CRM特性值来说，无论在计量学上还是在经济上都是至关重要的。

对CRM的使用范围也应加以说明。大多数情况下，使用范围已暗含在项目定义中，但是有时需要给出进一步的详细描述。这样的使用范围未必排斥其他用途，但是应该记住这样的用途不一定包含在所提供的证书或文件中。RM的使用范围可以依据法规和/或国际协议。

5.3 运输

在开始实际工作之前，着重要考虑的是一旦CRM研制完成，其运输过程是否满足现有法规的要求。许多CRM在人体直接接触时存在健康和安全的风险。合适的包装和正确的标签是满足运输（潜在）危险品法规的基本要求。有时，法律和法规禁止运输具有某些特性（如病毒、疾病）的材料，这意味着这种CRM根本不能被销售。因此，在开展实际定值项目之前，建议全面地审查运输和包装环节。

5.4 原材料收集

定值项目的首要工作就是获取足够量具有所需特性的原始材料。就基体材料而言，应注意可能存在材料特性的局限性。一些材料/特性的结合是罕见的，或者与其它特性的结合是罕见的，常常需要寻求一个折衷的方案。有时混合和/或加料技术可以解决这个问题。

所需材料的数量取决于下列因素：

- 所需 CRM/RM 样品的数量；
- 可行性研究的需要；
- 均匀性研究所需样品的数量；
- 稳定性研究所需样品的数量；
- 候选 CRM 测定所需样品的数量；
- 一次测量所需材料的量。

候选CRM所需样品的数量是一个商业问题，事先应仔细策划。重要的变量在于CRM使用寿命内将要发放的样品数量。由于使用寿命是内在稳定性的函数，这个变量也影响所需原材料量。例如，许多微生物材料由于有限的内在稳定性，因此预期其使用寿命比痕量元素干沉积物标准物质/标准样品短。如果每年发送同样数量的样品，微生物材料所需样品的量要比干沉积物的少。但对于微生物CRM，在第一年或在整个使用寿命内可能需要比干沉积物更多的样品用于稳定性研究。

5.5 可行性研究

当对生产和测定一个足够均匀和稳定的CRM的可行性有疑问时，可考虑开展可行性研究（见参考文献[11]）。诸如制备样品的最好方法、材料的稳定性或者与目标的吻合性等有关问题可在项目可行性研究中进行论证（见参考文献[11]、[12]）。有时也可组织进行可行性研究，以便参加测定的实验室对设备及其程序进行完善。对于以测定为目的的可行性研究，建议使用一批与候选CRM材料略微不同的材料。

5.6 要求的使用寿命和有效期

标准物质/标准样品的预期使用寿命是定值项目策划工作中的一个重要变量，另一个与稳定性有关的参数是CRM的有效期。根据影响材料稳定性机理的性质，可以采取各种措施延长有效期和/或使用寿命。控制水的作用是可考虑的首选方法之一，因为过分干燥或者过高的水含量都可能使材料不稳定。许多情况下，水分在导致基体和/或参数不稳定性的机理方面起着非常重要的作用。另外，为了抑制细菌活动，可以考虑对材料进行灭菌或消毒，但这些措施也可能对稳定性有负作用。有关稳定性和贮存条件的信息可以在文献中查到，也可以从相类似材料的用户处获取（例如工厂）。制备溶液时，添加剂可以增加样品的使用寿命和/或有效期。标准物质/标准样品的有效期既是贮存条件的函数，也是稳定性研究工作质量的函数，后者决定有效期可以外推到何种程度（见8.5）。

5.7 样品制备

5.7.1 概述

给出制备标准物质/标准样品的通用指南是非常困难的，本条款力图对某些特定问题提供指导，而并不仅仅是收集那些常常与定值项目成功高度相关、需要仔细考虑的一些问题。

5.7.2 人工合成材料

人工合成标准物质/标准样品，如纯物质，溶液和混合气体，采用与大多数基体标准物质/标准样品完全不同的方式进行制备。制备纯物质必须采用纯化技术减少杂质的总量，纯化技术的选择取决于目标主成分，可以包括蒸馏和/或重结晶技术。当制备一批CRM时，在分装之后，应按5.7至5.9的规定进行处理。

许多溶液和混合气体由于采用称量法制备，因此，通常能够很好地建立不确定度预算。与不确定度一样，原材料的纯度（或组成）也进入计算候选CRM组成的模型。批量材料制备也广泛使用容量技术，通常容量法稍易于进行，但与称量法制备相比其不确定度较大。

5.7.3 混合

如果认为某个特性值太高或太低，可以考虑将两种或多种基体混合，该过程最好采用相似种类的基体进行，尽管“相似种类”仍可能存在很大的差异。为了正确地混合，应对材料中颗粒结块进行控制。材料中水分含量常常是决定因素，如果材料是“风干”的，在“好”的混合过程中一般（但当然不总是）不会出现结块。类似泥浆的材料就是这样。混合时如果结块不消失则可能存在问题。事实上颗粒结块是不可避免的，例如，含水量小于2%的大豆粉仍是粘性的。

使不同基体正确地混合还要求被混合材料的密度和颗粒粒径分布充分相似，而且分布应足够窄，这实际上将减少偏析风险。采用合适的工艺及正确运用颗粒粉碎和混合技术，一般可获得均匀性和稳定性均良好的材料。

如有所怀疑，可以对已混合的材料进行一次快速均匀性检验，即从混合好的材料中取出几份来研究待测特性的均匀性。这种检验可以只做少量几份，但如要获得均匀性的一些信息则要多做一些。一般认为10份就能提供有意义的结果，可以决定诸如混合材料是否适于进一步加工之类的问题。

5.7.4 加料

在某些情况下，加料可作为标准物质/标准样品生产的一种合适方法。例如，由固体材料制备的提取物；再如猪脂肪中三个PCBs系列CRM，该CRM在高温时是液体；加料法制备预定特性CRM的良好例子还有液体、金属和合金、油和工作场所空气。

加料的主要问题是要使候选标准物质/标准样品具有足够的均匀性和稳定性。采用适当的加料方法可以使材料、甚至是固体材料满足均匀性和稳定性的要求。例如，固体加料的一种合适方法是“刚好润湿”技术，即将加料成分溶于适量溶剂中、使其刚好能足以完全润湿固体表面。在选择溶剂时，应选择蒸发速度是可控制的溶剂。蒸发速度太快，加料物可能会再从微孔和微束中出来，不能很好地与表面结合，影响材料的稳定性。蒸发速度太慢将会使基体中原有的其它组分溶出、甚至损失。

但是，对某些基体CRM来说，加料肯定不是一个获得具有待测特性期望值材料的合适方法，因为它可能使得CRM与日常样品的性质完全不同。通常可以预期天然的与加料的待测物主要差别是在结合上，这会产生差异如萃取行为。因此，应检查加料的材料和自然（污染）材料间等效性，确保材料可代表实际样品。

5.7.5 均匀化与分装

采集的材料通常需要经过几个制备步骤才能成为标准物质/标准样品。其中必要的步骤包括干燥、粉碎、筛分、稳定化和分装/装瓶。在项目设计时，应确定样品制备将要达到的程度。例如，可以将采集的材料制备成能够直接测量的一种提取物。但是多数情况下，制备样品时最好保持所采集材料的原有状态，尽管样品制备过程通常会使其不均匀性减小、稳定性增加。

RM特性值所要求的不确定度和使用寿命决定了对样品制备技术选择的要求。应牢记候选标准物质/标准样品的制备方式将影响其可能的应用，例如，分发提取物将使其不可能检查客户实验室提取步骤的准确性。因此，在决定如何将原料制备成符合应用范围、适于被定值的CRM时，应牢记CRM的制备目标。

5.8 均匀性研究

在批定值项目中必须进行均匀性研究，以证明一批瓶（单元）是足够均匀的。质量保证与测定批的瓶间变差一样重要，后者是CRM特性值不确定度估计值中应包含的一个不确定度分量。甚至当预期材料与溶液一样均匀时，瓶间不均匀性评估也是需要的。当处理固体标准物质/标准样品时，包括泥浆和淤泥，应通过瓶内均匀性研究来确定最小取样量。原则上该均匀性研究不会增加有关特性值的不确定度。额外所需样

品的数量主要取决于瓶间均匀性研究，随机抽取的最小瓶数在10和30之间，而且一般不应少于10。

进行均匀性研究的最合适样品数量可根据统计设计技术进行确定，设计时通常需要考虑例如由于测量不确定度而不能检测出任何不均匀性的情况，而且，瓶数取决于批的大小，因此，从一批中抽取的样品数可认为是整批的代表。这个要求应与测量不确定度平衡，后者是（在重复性条件下）测量的重复性标准偏差和重复测定次数的函数。以上所提到的统计技术可能有助于平衡瓶数和重复测定次数，以便选择最佳方法。

5.9 稳定性研究

稳定性检验的目的是确定候选RM在制备后的不稳定性或材料的稳定性，即使“稳定的”材料也可能会有一个或多个特性值显示不稳定。稳定性可分为：

- 在规定贮存条件下的稳定性（长期稳定性）和
- 在规定运输条件下的稳定性。

与均匀性研究一样，质量保证与确定不稳定性引起的不确定度预算同样重要。长期稳定性是有关CRM特性值在指定贮存条件下仍存在的不稳定性。因此，重要的是规定这些条件并研究在该条件下材料的稳定性。参考温度的选择应确保材料在该温度下是稳定的。尽管确定了最佳贮存条件，但许多生物和环境标准物质/标准样品在某种程度上仍表现出不稳定性。同样应选择运输条件以使材料在运输中的不稳定性不超过其在生产者库房的不稳定性。因此，短期稳定性是在贮存条件之外指定运输条件（从生产者到使用者）下影响CRM稳定性的一个不确定度分量。

短期稳定性研究通常在不同温度下进行，以研究不同温度条件对材料特性的影响。运输中样品温度可能在 -50°C ~ $+70^{\circ}\text{C}$ 范围内变化，这取决于包装类型和运输方式。根据观察，可以确定运输条件并拟定包装说明，以便有效地消除不希望出现的副作用。短期稳定性研究一般要进行一至二个月，当同时要确定最佳贮存条件时，则可以延长。

稳定性研究需要相当多的瓶（单元）数。在每个时间点，应使用一瓶以上。由于大多数长期稳定性研究持续24到36个月，一般5到6个时间点，因此在每个温度下至少需要10到12瓶。当设计预定多个温度时，瓶数应相应增加。短期稳定性研究通常使用三到五个时间点，超过两周，按照与长期稳定性一样的推算，短期稳定性研究在每个温度下应需要6到10瓶。材料的不均匀性也将影响稳定性检验所需的单元数。如果材料是非均匀的，对多瓶进行单次测定比重复测定较少几瓶更好。

批定值稳定性研究的最好方法是在重复性条件下工作，否则，由于稳定性检验中因再现性对结果的影响会使得不稳定性产生的估计不确定度不必要地放大。使用同步设计（见参考文献[13]）使得在重复性条件下工作成为可能。所有样品都保存在同一参考温度，并假定在该温度下不会出现不稳定现象（不一定是预设的贮存温度）。将样品置于稳定性研究的试验温度至所有样品测量完。时间点根据从样品被置于试验温

度那一刻到样品被测量那一刻之间的时间间隔确定。

在经典方法（见8.2）中，应选择具有良好再现性的测量方法。由于保持测量方法良好的再现性比单独一轮测量中保持良好的重复性要难得多，因此，同步设计优于经典设计。除此以外，采用经典设计评估的不确定度总比同步法大，这意味着通过同步稳定性研究得到的有效期（在给定的不确定度水平下）比经典法要长一些。这些优点很好补偿了其在稳定性研究期间没有数据的缺点，特别是那些重复性和再现性相对较差的方法。当需要中间数据时，则应独立于同步稳定性研究进行测量。当单个制品定值时，只能选择经典方法。

稳定性研究的实验设计，包括测定最佳的时间点数和样品数，当预计到测量方法不能检测任何不稳定性时可以以统计设计为基础，而且，由于经验模型用于大部分稳定性研究，因此，时间点的数量应足够大到对模型有效性进行正确评估。例如，有两个参数（截距和斜率）的线性模型至少需要3或4个点，但为了更准确地评估通常要多做些点。应用多参数模型进行稳定性研究时，时间点数应相应增加。

5.10 测量方法选择

用于均匀性研究的测量方法应该具有良好的重复性和选择性。对于不在同一天进行样品测量的稳定性研究来说，测量方法的选择性和再现性最重要。因此，用于均匀性和稳定性研究的测量方法不一定相同。只要均匀性、稳定性研究以及测定的结果建立了与共同测量标准的溯源性就不会有问题。这个测量标准可以是一种材料，其适于评估不同的校准或由不同测量方法获得的结果。确保定值项目中所有测量的溯源性是一个重要的要求（见GB/T 15000.7—2001和参考文献[14]）。

候选标准物质/标准样品、特别是基体标准物质/标准样品的测定，通常最好采用多种方法和多个实验室定值。方法和实验室能力都应该代表“最高水平”，并且它们应能使其测量溯源到项目设计规定的测量标准。

候选标准物质/标准样品的测定可以采用不同的方式，其中有两个主要方式：

- a) 采用单一方法测定
- b) 采用多种方法和/或多个实验室测定。

方式b) 包括被称为协作研究或协作试验的实验计划。这两个名称强调了协作者和参加者测定RM的共同努力。总之，测定中使用的所有测量程序均应可溯源到“规定测量标准”，最好溯源到SI单位。测量结果溯源性涉及面远远超出实际测量，它还包包括样品的转化。样品转化是指将材料（或制品）从一种（物理的、化学的）状态转变为另一种状态。例如，样品转化包括破坏样品和提取被测物。

最后，为了得到特性值及其标准不确定度的正确估计值，应将均匀性研究、稳定性研究和材料测定的测量结合起来。对数据的一个要求是其应参照相同的“标度”，即所有的测量都是由经过正确校准的设备完成，并且这些校准结果之间相互可比。特别当参加实验室较多时，可以使用某种标准纯物质、混合物和溶液来验证校准之间的

一致性。这既是确定适当标准从而建立计量溯源性的问题，也是能够证明项目各个阶段所获得结果的有效性和可比性的问题。

5.11 证书

CRM的证书要求见第11章规定。

5.12 项目设计小结

总之，标准物质/标准样品制备包括以下步骤：

- a) 标准物质/标准样品的定义，例如：基体，待定特性及其预期的浓度水平和不确定度水平；
- b) 采样程序设计；
- c) 样品制备程序设计；
- d) 适于均匀性和稳定性检验的测量方法的选择；
- e) 标准物质/标准样品测定的设计；
- f) 采样；
- g) 样品制备；
- h) 合适的测定方法的选择；
- i) 均匀性检验；
- j) 稳定性检验；
- k) 标准物质/标准样品的测定；
- l) 均匀性检验、稳定性检验和测定结果的总结，包括测量不确定度的全面评估；
- m) 证书设计和定值报告（如果适用）。

6 测量不确定度评估

6.1 CRM/RM 特性值不确定度评估基础

GUM是所有测量不确定度评估的基础。大多数CRM都可以通过GUM:1993第8章的方法进行评估。CRM不确定度的评估程序概述如下：

- a) 用数学式表示待定特性值与所有输入量之间的关系，其中应包括对特性值不确定度有显著贡献的所有量，这种关系式称作测量数学模型。
- b) 根据数据列的统计分析或其它方法确定所有输入量的值。
- c) 评估所有输入量的标准不确定度：由数据统计分析获得的量采用 A 类评估，其他的量采用 B 类评估。
- d) 求出各输入量之间的协方差。
- e) 计算特性值(x)，即待定特性值。
- f) 应用 GUM:1993 第 5 章中给出的传播率，由输入量的标准不确定度和协方差确定特性值的合成标准不确定度。

- g) 确定包含因子 k 以获得扩展不确定度 U 。为此可假设区间 $[x - U, x + U]$ 能够合理地包含待特性值分布的大部分。包含因子应基于需要的置信水平（通常为 95%）、 y 的概率密度函数以及自由度（如适用）来选择。
- h) 根据 GB/T 15000.4—2003 的要求，报告特性值的同时应报告扩展不确定度 U 和包含因子 k 。

绝大多数情况下可采用上述方法。然而正如 GUM 中讨论的那样，在某些情况下，应该选择其它一些方法，这包括：

- 没有严密的数学模型描述特性值与输入量之间关系，以及
- 应用不确定度合成和传播率得到的线性近似显然是无效的。

此时，可采用其它统计技术，包括蒙特卡洛 (Monte Carlo) 法或自举 (bootstrap) 法，确定 CRM 特性值的不确定度。本指南假设可采用上述方法，其它情况应按 GUM 规定处理。

如何评估单个不确定度分量在 GUM 中有详述。在许多情况下，为了利用现有数据，如参考文献 [15] 第七章中所描述的确证研究结果，可以规定一些组合的不确定度分量。

对包括批不均匀性、长期稳定性和短期稳定性引起的不确定度分量进行评估需要专门指导。本指南第 7 章（均匀性研究）和第 8 章（稳定性研究）包含了这些不确定度分量的评估，第 9 章和第 10 章中给出了批测定的测量不确定度评估的一些附加指导。

基于单点校准定值的单个制品的 CRM 特性值的不确定度，可使用 GUM 中所述的正常程序来评估，但应注意此类 CRM 不确定度的预算也应包括长期稳定性的影响。

6.2 批测定的基本模式

建立测定过程的不确定度评估模型既不是常规任务也不是严格的数学任务。为特定候选的 CRM 特性值建立一个合适的模型是一项复杂的任务，这需要仔细地考虑到该标准物质/标准样品生产和定值程序中所有有关的细节。一个基本要求是模型中应包括能够对 CRM 特性值的不确定度有显著贡献的所有因素。为了保证标准物质/标准样品的合成不确定度的完整性，应注意除了批测定之外，均匀性、长期稳定性和短期稳定性也对标准物质/标准样品的合成不确定度起着重要作用。因此，标准物质/标准样品的不确定度可表示如下：

- a) 由批（测定）得到的标准值的不确定度；
- b) 转换成单个包装（均匀性）；
- c) 分发到消费者（短期稳定性）；
- d) 销售的时间（长期稳定性）。

CRM 特性值不确定度的这种定义考虑如下因素：

- e) 材料的严密测定伴随有不确定度；
- f) 用户一般每次只用一个样品；

- g) 材料将被生产者/销售者储存较长的一段时间；
h) 材料必须运输到用户。

上述因素对CRM被测量的给定值（也就是标准值）的不确定度都有显著贡献。对这些影响因素的鉴别不能超出常规的操作，不确定度评估不宜用来也不应用来解释CRM的事故、错误、不恰当使用、不恰当运输等等。这种方法与GUM:1993, 3.4.8相一致。

该模型可表示如下：

$$x_{CRM} = x_{char} + \delta x_{bb} + \delta x_{lts} + \delta x_{sts} \quad \dots\dots\dots (1)$$

式中：

- x_{CRM} ——特性值；
 x_{char} ——由批测定得到的特性值，或在测定单个制品时，由该制品得到的特性值；
 δx_{bb} ——瓶间变差的误差项；
 δx_{lts} ——长期不稳定性引起的误差项；
 δx_{sts} ——短期不稳定性引起的误差项。

通常，做均匀性和稳定性研究时这些误差项设计为零，但是其不确定度并不为零。假定各变量是独立的，与CRM特性值相关的的不确定度可表示为：

$$u_{CRM} = \sqrt{u_{char}^2 + u_{bb}^2 + u_{lts}^2 + u_{sts}^2} \quad \dots\dots\dots (2)$$

使用GUM:1993, E.8中的不确定度传播律，不确定度分量 u_{bb} （瓶间标准不确定度）， u_{lts} （长期稳定性标准不确定度）和 u_{sts} （短期稳定性标准不确定度）对应于模型中的误差项。与CRM特性值相关的合成标准不确定度也可能与材料的有效期（见第8章）有关。

长期稳定性有时是时间的函数，如放射性同位素标准物质/标准样品，在评估此类CRM特性值不确定度时所用的模型应考虑标准值与时间的相关性。

在GUM和一些补充文件如参考文献[15]中给出了建立测量模型的通用性建议。第7章到第10章针对标准物质/标准样品定值的关键步骤给出了建模和数据评估的一些专门指导。

有时可能会偏离上述的基本模型，如样品不需要经过运输，或者明确说明说明书中的不确定度不包括样品运输带来的不确定度。第7章（均匀性研究）、第8章（稳定性研究）以及第9和10章（特性值的测定）中给出了评估这些不确定度分量的专门指导。

例如，一个对温度升高敏感的样品，生产者预计将其从生产者运送到消费者最多只需1个星期，而实际用了6个星期，此时，CRM的特性可能已严重变质。如果生产者在证书或其附件中做了详细说明，生产者就有权相应地限定这种短期稳定性研究的条件。

6.3 不确定度来源

除了已提到的不确定度来源之外，测量程序中普遍存在的不确定度来源也应包括在模型中。在GUM和参考文献[15]中均列出了这些不确定度来源，建议在建立模型时采用该通用清单，以减少获得所有相关不确定度分量的工作量。通常，测量方法已经依据测量不确定度进行了评估，适用于这些方法的模型也可用于评估CRM特性值的不确定度。需要注意的是，一个特定测量程序中的任何改变随之都应重新评估其不确定度模型。

测量方法的不确定度模型通常包含一些组合的不确定度分量，即依赖于其它几个分量的不确定度分量。即使这些组合的分量在测量方法用于日常测量时不显现，也可能导致协方差。协方差和相关性的评估对于获得CRM特性值的合成标准不确定度的正确估计至关重要。为方便检出协方差，建议将包含在组合的不确定度分量中的不确定度分量制成文件，该文件允许较快地鉴别出协方差和相关性的可能来源。GUM:1993的附录F给出了如何评估协方差的进一步指导。

6.4 分布函数问题

大多数统计方法需要对所研究变量的概率密度函数做出隐含或明确的假设。由于在评估过程中要对概率密度函数进行设定或假设，因此GUM的方法也不例外。标准物质/标准样品的定值模型由于建立在基本统计学理论基础之上，因此也不违背这个原则。这些假设通常都是隐含的、而不是明确的（如通过使用特殊统计估计/B类评估方法），例如，许多统计方法假定数据呈正态分布。本指南中的大多数统计也是基于此假设。对于成分测量的多数数据这种近似是合理的，而对于其它一些测量，如少量数据的计算，该假定可能是无效。

回归分析和方差分析都是基于数据呈正态分布的假定。然而，只要是用于估计方差，这些统计工具都能很好地处理呈单峰分布的数据（例如，见第7章、第8章和参考文献[21]）。

当测定数据的实验分布呈现多个极大值时会产生严重的问题。最糟糕的是由于实验室间和/或方法间结果的不一致性导致材料不能被定值。只有在方法间或实验室间结果一致时才能确定单个特性值。微小的差异可以通过引入附加的不确定度分量来解决。如果使用同一种特定方法的实验室间的结果一致，那么可以考虑依赖于方法的定值，结果得到依赖方法的特性值；如果实验室间结果不一致并且按方法分组也解决不了这问题，那么测定数据就不适宜确定特性值。

6.5 比率的运用

使用比率时可能会有问题，如在稳定性研究中^[16]。应注意两个呈正态分布的变量的比率不一定也呈正态分布^[17]。两个变量比率的实际分布不仅依赖于所涉及的两个变量本身的分布，而且还依赖于这些分布函数的参数的实际值。特别是当特性值可能为零时，如进行某些颜色测量，由于比率遵循柯西（Cauchy）分布，因此，使用比率可

能引起问题。柯西分布没有矩量，这意味着方差没被定义，故不可能在指定概率分布的基础上评估测量不确定度。

通过正确运用不确定度传播律可以使测量带来的随机误差分量最小化，即在形成比率的两个变量间引入必要的协方差项。这种不确定度的降低常常是使用比率^[16]想要达到的效果。对于制品的比率分布函数而言，对观测数据运用不确定度传播律具有安全的优点，同时也能达到“抑制随机误差分量”的效果。

6.6 包含因子选择

6.1程序g)中使用的包含因子是基于特性值的分布函数（通常为正态分布）和置信水平（通常为95%）而确定的。因此，在此基础（正态分布，95%的置信水平）上包含因子常取 $k=2$ 。当认为（有效）自由度数很低时，可以用学生 t 分布替代包含因子。

在特性值的分布不对称的情况下，如按泊松（Poisson）分布计算结果，则应以置信区间表述而不是用扩展不确定度和包含因子。

6.7 重新定值

当超出规定的时间时，CRM实际的特性值可能会偏离标准值。当CRM实际的特性值超出证书规定的范围以外时，有两种基本的处理方法：

- 收回 CRM/RM；
- 对标准物质/标准样品重新定值。

选择哪种方法取决于经济和技术两个方面的因素。收回的技术方面的原因可以是诸如基体变质或一个、多个组分变质，这可以通过稳定性检验或稳定性监控（见8.4）查出。当剩余的CRM数量少到不值得重新定值时，从市场上收回CRM也许是较好的选择。

重新定值意味着标准物质/标准样品的均匀性检验、稳定性检验和/或测定等相关的工作需要重新进行。专业领域测量能力的提高也可能是重新定值（经济上可行时）的原因。如果所生产的材料还足够好，那么，减小特性值的不确定度，提高特性值的定值质量，可以使得CRM依然对使用者有用。

实际工作中还有一类重新定值的原因是由于标准物质/标准样品的特性逐步发生了变化。此类实例之一是煤的热值，即使在最好的条件下保存，随着时间的推移，该值也会发生变化。

7 均匀性研究

7.1 概述

大多数RM是按单元（例如瓶、小瓶或试片）成批制备的，许多RM制备的最后一步是将之分成可使用的单元。按照抽样方案从批单元中选取一个子集，通常由10到30个单元组成，进行均匀性研究。选取子集有许多不同的方法（例如，随机抽样、分层随机抽样或系统抽样），在实践中大多采用随机抽样或分层随机抽样方法，这两种方

法通常可以提供一个被认为能代表整批的子集。但当肯定不能检出批不均匀性时，也可采用系统抽样方案。

瓶间均匀性研究³⁾的结果可用于评估定值模型中的一个不确定度分量(见第6章)，该不确定度分量的大小可能变化很大，主要取决于RM的性质。此类均匀性检验只适用于批量生产。

第二种重要的不均匀性类型是瓶内均匀性，通过提供正确的使用说明可以明显地降低其影响。使用说明可以包括样品的再混合和针对颗粒物而言的最小取样量，该取样量是指当正确取用时，在定值不确定度范围内能代表RM的最小检验用量。

7.2 材料

以溶液或纯化合物（如果被定值的是纯度，而不是杂质）形式制备的RM，从物理（热力学）的角度来看具有高度的均匀性，但这些材料也会由于密度梯度或金属中气孔而表现出某些不均匀性。对这些材料进行均匀性检验的目的主要是为了检测制备期间可能由于未被发现的问题所产生的杂质、干扰物或异常。这些情况下，即便不能忽略，预计从瓶间均匀性研究得到的不确定度贡献也很小。即使在可以被假定为理想的均匀性情况下，这种假定也应通过均匀性研究的试验加以验证。

粉末混合物、矿石、合金等材料由于本质上其组分是不均匀的，因此，由这些材料制备的RM应通过检验来评估其不均匀度。与稳定性检验或测定的不确定度相比，瓶间不均匀性带来的不确定度分量的大小仍可能很小甚至可以忽略。但有些情况下，不可避免的是它与特性值测定的不确定度分量的大小相同。多数情况要依靠制备过程中可利用的选择方案来降低批不均匀性。

7.3 均匀性的概念

理论上，如果某给定特性值在材料的各部分间（或单元间）没有差异，那么该材料对该特性而言是完全均匀的。但是实际上，当材料的各部分间（或单元间）某给定特性值的差异与测定的不确定度分量相比可以忽略不计时，可以认为该材料对该特性而言是均匀的。

检出批不均匀性(u_{bb})是有试验方面限制的，应注意不要低估来自诸如测量方法的限制等引起的不确定度分量。特别是当仅有重复性不好的方法可用时，会存在低估的风险。此外，在可能的情况下，抽取用于测量的子样数要足够多，以确保抽样不会对测量重复性不确定度有显著贡献（见7.10）。

本章主要涉及批不均匀性，大多数标准物质/标准样品最为关注的也是此类不均匀性，不过也有些重要的例外。例如，在表面分析中，标准物质/标准样品可以是薄片或箔，关注的是沿着表面方向的不均匀性而不是垂直于表面方向的不均匀性。本章中考虑的大多数问题对其他类型的不均匀性也有效，但提供的指导应与更多的专业文献进行比较，包括描述相关测量方法的国际标准。

3) 此处文献采用“瓶间”均匀性，可以理解为同样也可以用于其它物理形状的RM，如小瓶或测试片。

7.4 实际应用

最理想的是应该能表征RM的每个待定特性的不均匀度。对于待定特性相对较多的RM，评估其全部特性的不均匀度无论是经济上还是实际工作上也许都是难于负担的，并且在某些情况下也是不可行的。不过应认识到所研制RM的质量首先取决于对批不均匀性的正确评估。

在实际工作中当理想的方法不可行时，RM的均匀度（在某些条件下）可以仅就所选择的特性进行评估，但应根据已确立的化学或物理关系适当地选择这些特性。例如，RM矿物相中元素间的伴生使之能合理地假定该RM对未选择元素也有类似的均匀性；均匀性研究中可以缩减特性的例子还包括无烟煤和褐煤，其不均匀性可以适当地通过灰份量和/或含硫量来反映，对其它特性而言，批均匀性通常好于这些特性的均匀性。

在所有的情况下，对于均匀性研究实验未涉及特性的均匀性应获取附加的证据，该证据可以从文献、稳定性研究或材料的测定中获得。在有充分证据表明不均匀性度未被低估时，如此获得的证据可以允许将一个特性的不均匀性大小定量转移到另外一个特性之上。

7.5 测量

均匀性研究中的测量应在重复性条件下进行（重复性条件的定义见GB/T 6379.1—2004，且测量方法的重复性标准偏差应很小。如果可能，应使单瓶测定的不确定度 (s_r/\sqrt{n}) 显著小于（预计的）特性值测定的合成标准不确定度。但有时这是不可行的，需要使用7.9给出的替代方法^[19]，不过这样一般会导致更大的不确定度估计值。

测量应采取可以将测量趋势（漂移）与样品趋势分离的方式进行。这可以通过按随机顺序重复测量均匀性研究的样品达到。或按如下示例在重复测量之间颠倒测量样品的顺序：

示例 假定有 10 个用于均匀性研究的样品，每个重复测量 3 次。适合的测量方案如下：

第一次：1-3-5-7-9-2-4-6-8-10

第二次：10-9-8-7-6-5-4-3-2-1

第三次：2-4-6-8-10-1-3-5-7-9

对正确测量顺序的结果作趋势分析可以发现测量中的漂移趋势。将瓶的平均值作为其序号的函数来分析可以发现样品制备中的趋势。因此，重要的是一批样品的序号与样品的制备过程，特别是分装过程的逻辑关系。

7.6 统计上有效的抽样方案和趋势分析

在均匀性研究中抽取瓶（单元）的抽样方案可以是随机的、随机分层的、有时也可以是系统的。抽样方案应考虑样品制备方法的潜在缺点，对制备的批料进行严格检查。随机分层抽样由于可以保证用于均匀性研究所抽取的瓶均匀分布于整批中，因此，

在许多情况下推荐采用之。当实践中不存在批内系统影响或趋势的风险时可采用系统抽样方案。

测量应以这种方式进行，确保样品中可能出现的任何趋势都不干扰测量本身可能的任何趋势。例如，在测量方案中可通过将瓶的顺序随机化并结合改变样品的测量顺序达到此目的。

在确定瓶间均匀性标准不确定度的大小之前，要检查所得到的实验数据是否有趋势。在8.3.1中，稳定性研究部分已给出了趋势分析的基本方法，该方法也可用于以瓶号为函数的均匀性研究数据。如果瓶间存在显著趋势，那么通常生产的这批材料不适合批定值。测量结果中的趋势无论统计上是否显著都要作某些校正。趋势分析和校正仪器漂移的方法都包含能直接用于仪器的质量控制样品。当发现批内有某种趋势时，有必要重新设计分装程序以有效地消除这种趋势。

7.7 均匀性研究评估

均匀性研究的基本模型包含 $i = 1 \dots a$ 瓶和 $j = 1 \dots n_i$ 测量次数，模型表示如下（见参考文献[20]到[22]）：

$$x_{ij} = \mu + A_i + \varepsilon_{ij} \quad \dots\dots\dots (3)$$

式中：

x_{ij} ——均匀性研究中单次测量结果；

μ —— x_{ij} 的（数学）期望值，即当重复测量的次数趋于无穷大时所得到的总平均值（平均值的平均值），如果测量没有偏移，那么 μ 就等于真值；

A_i ——瓶间均匀性的误差项；

ε_{ij} ——随机测量误差项。

A_i 和 ε_{ij} 的方差分别是瓶间方差和重复性方差，通常可假定 A_i 和 ε_{ij} 是相互独立的，即瓶间不均匀性不影响测量的重复性，反之亦然。此外，通常还可假定变量 A_i 符合均值为零、方差为 σ_A^2 的正态分布，同样也可假定随机测量误差 ε_{ij} 符合均值为零、方差为 σ^2 [21] 的正态分布。

瓶间均匀性研究可以采用不同的实验设计，B.2描述了一个采用全嵌套单因素方差分析法的实例。

7.8 瓶间均匀性研究

瓶间均匀性研究的目的是测定瓶间的变差。如前面的条款所述，“组”代表瓶（单元）。图1和图2示出两种典型的瓶间均匀性研究的实验方案。

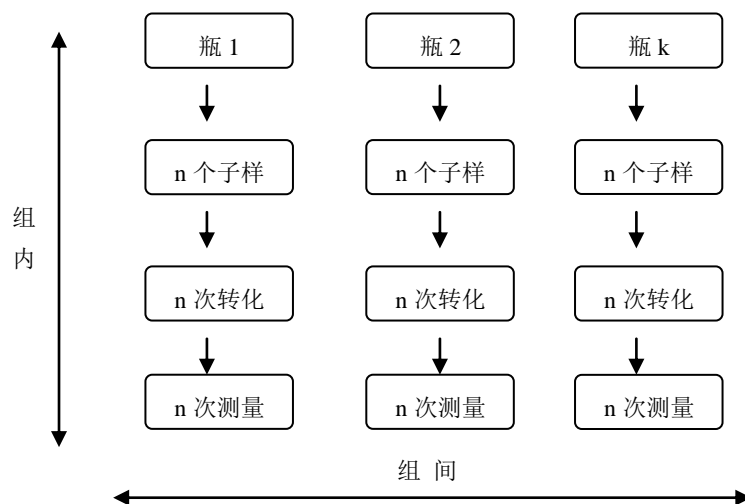


图1 瓶间均匀性研究设计图

(引自参考文献[22])

图1为各单元可以进行二次取样并完成二次取样的理想情况。在该设计中，由于从该批的每个样品中都抽取了多份检验用样并单独进行了转化，因此瓶间方差仅包括瓶间不均匀性，而瓶内方差却包括来源于测量、转化和二次取样的不确定度。从获得材料不均匀性的无偏估计的角度来看，这是一种理想的情况。

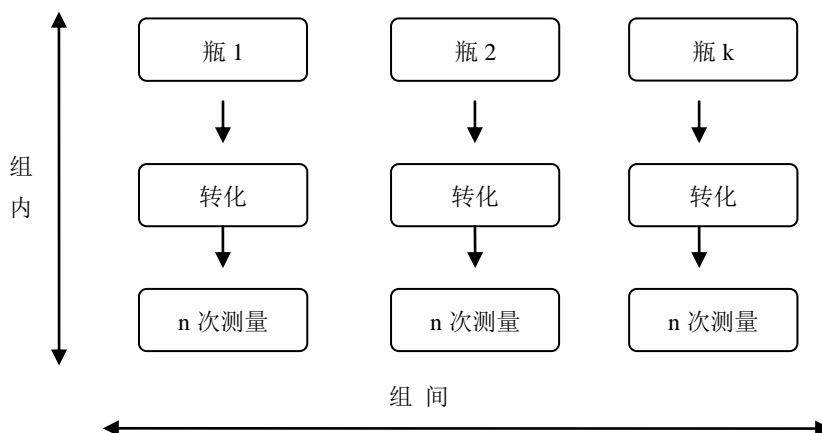


图2 瓶间均匀性研究的替补设计图

(引自参考文献[22])

图2为单元不可以进行二次取样或诸如因经济等原因而没有完成二次取样的情况。在该设计中，由于瓶间均匀性的影响来源于样品的转化，所以它被包括在“组间”方差中。“组内”方差只包括测量的重复性。试片或“一次性”样品由于通常只能测试一次，重复测定次数 n 等于1，因此没有必要考虑瓶内均匀性的影响；样品转化后可以多次测量时， n 通常会较大。当 $n > 1$ 时，数据可以用方差分析来处理（见附录A.1和附录B.2）。

如果采用单因素方差分析法，那么 s_{bb} 在两种情况下都可通过下式计算。

$$s_{bb}^2 = s_A^2 = \frac{MS_{among} - MS_{within}}{n_0} \dots\dots\dots (4)$$

此时瓶间方差 s_{bb}^2 等同于 u_{bb}^2 。

7.9 测量方法重复性欠佳的情况

由于使用可充分重复的测量方法进行均匀性研究未必总是可能的，因此，需要一种替代方法估计出最大影响。如果 u_{bb} 表示来源于定值模型中批不均匀性的不确定度分量，则应注意的是：

$$\frac{MS_{among} - MS_{within}}{n} \leq u_{bb}^2 \leq s_{bb}^2 + \frac{s_r^2}{n_0} \dots\dots\dots (5)$$

重复性方差可以单独导出，或者设定其等于 MS_{within} 。表达式右边表示一瓶测量结果的标准不确定度的平方，表达式左边表示“纯”来源于方差分析估计得出的瓶间不均匀性的影响。

参考文献[19]讨论了测量方法重复性欠佳时替代公式（4）的结果进行不确定度估计的各种方法，重复性标准偏差对 s_{bb} 的影响可按式计算：

$$u_{bb} = \sqrt{\frac{MS_{within}}{n}} \sqrt{\frac{2}{v_{MS_{within}}}} \dots\dots\dots (6)$$

式中：

MS_{within} ——瓶间均匀性研究中测量的重复性方差。

这个表达式是基于以下考虑：可为 s_{bb} 设定置信区间，以及95%置信区间的半宽（转化为标准不确定度）可以用来衡量方法重复性对 s_{bb} 估计值的影响。该表达式是一个如何解决不均匀性不能被估计的情况例子。可以选择其它方法，但应满足公式（5）的判据。

7.10 瓶内均匀性

只有当候选RM（单元）可以被二次取样时，才会有瓶内均匀性的问题。在许多情况下，不可能获得瓶内不均匀性引起的方差的准确估计值。在某种程度上，瓶内均匀性的估计值中总包含检测方法的重复性，这使得 s_{wb} 的估计值总是“安全的”，即比实际不确定度大。图3为瓶内均匀性研究的设计图。

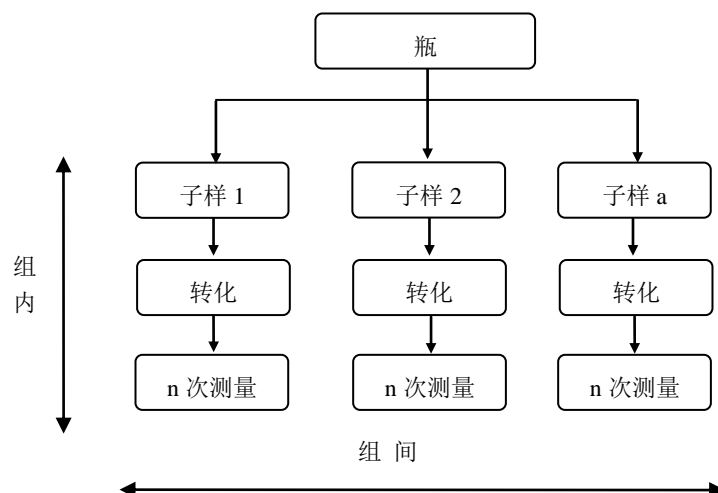


图3 瓶内均匀性研究的设计图

(引自参考文献[22])

从一个样品中取出的多份检验用样通常只能被转化一次（图3），但也有同一份检验用样可以被多次测量的例外情况（如应用X-荧光光谱法）。如同瓶间均匀性的情况（见附录A.1），此时，也可以考虑采用单因素方差分析法，相关的标准偏差是组间标准偏差，而一组则代表一个子样。

最小取样量通过不同检验用样量的瓶内均匀性研究确定。由于瓶内均匀性标准偏差取决于带有特性的粒子数量，因此确定最少的粒子数量（或最小检验用样量）是可能的。这个最小值是能满足检验用样的标准偏差等于测量方法重复性标准偏差的最小取样量。

最小取样量可通过实验或外推法来确定。通过不同取样量得到的瓶内标准偏差外推可以获得不影响特定参数测量重复性的最小取样量。由于瓶内均匀性标准偏差通常被高估，因此，最小取样量通常也会被高估。

另一种解决问题的方法是对于某特定的取样量证明其检验用样量的标准偏差等于测量方法的重复性标准偏差。这可以通过 χ^2 -检验比较方差（详见GB/T 15000.8—2003）进行评估，实验中的取样量可定为最小取样量。

8 稳定性研究

8.1 稳定性类型

标准物质/标准样品定值时，需考虑两种类型的（不）稳定性：

- 材料的长期稳定性（例如：有效期）；
- 材料的短期稳定性（例如：运输条件下材料的稳定性）。

标准物质/标准样品的长期稳定性与RM在生产者的贮存条件下的行为有关。短期稳定性与样品运输过程中的外部因素有关。有时，由于在运输中不可能保持标准物质

/标准样品稳定性所需的合适条件，因此应考虑在特性值中增加一些不确定度。

对于CRM证书中所述的不确定度的有效性而言，正确估计长期稳定性和短期稳定性的影响与正确估计批不均匀性（见第7章）同样重要。在CRM的使用寿命内，应对证书中不确定度的有效性予以证实，以便CRM能满足稳定性的需求。

若不能保持适宜的运输条件，了解样品可能发生的变化同样重要。许多情况下，在首次使用CRM之前对其进行简单的检验足以重新确认证书的有效性。而在另外一些情况下，CRM显然已经失效。这种认识可以使生产者能给出更好的建议，对使用者而言就是向其提供更好的产品。因此稳定性研究不仅指导了与材料稳定性相关的测量不确定度评估，而且也能确定适当的贮存和运输条件（见5.9）。

8.2 实验设计

稳定性研究有两种基本实验设计^{[13], [23]}：

- 经典稳定性研究；
- 同步稳定性研究。

在经典稳定性研究中，随着时间的流逝，对同一时间所制备的样品（即一批样品）在相同的条件下进行测量。此时，由于测量是在（实验室内）再现性条件下进行的，其中包括了测量系统的不稳定性，因此会导致相对较高的不确定度，

采用同步稳定性研究可以使所有稳定性研究的测量能在重复性条件下进行^[13]，即一次测量做一次校准。“同步”这个词强调所有测量同时进行，而不是像经典稳定性研究中那样分配在稳定性研究的时间间隔上。

同步方法由于减少了时间点上的离散性，因此改进了稳定性研究的“分辨力”。由于测量重复性和（实验室内）再现性之间的差异，使得同步稳定性研究通常比经典方法得到的不确定度要小。这种设计的先决条件是能规定一定条件，在此条件下不发生降解或至少以不同于选定的贮存条件下的速率发生降解。同步设计是专为批定值设计的，不适用于单个制品。

两种实验设计都适用于长期和短期稳定性研究。短期稳定性研究主要研究在推荐（指定）的包装和运输条件（最低要求）下材料及其特性值的行为。运输条件的限制越多，短期稳定性研究工作就越少。推荐用于运输的条件为材料的不稳定性不大于长期稳定性研究中的不稳定性，以便定值时不需要计入短期稳定性的不确定度贡献。对于几类标准物质/标准样品（例如：临床、生物和环境标准物质/标准样品），很明显不一定总能保持其运输条件，使运输对材料的影响可以忽略。在涉及特定的基体/特性组合而又没有以往的经验时，可在不同的温度下进行短期稳定性研究，以获得适当贮存条件以及运输中需采取预防措施的信息。

这种研究通常大约需要两个月，为获得长期稳定性的更多信息，也可延长至6~12个月。由于两个月后的研究只涉及贮存条件，此时温度范围可以适当缩小。CRM的运输能够并应该以这样的方式加以组织，即运输需要的时间应尽可能短。经验表明，

两周以内是可行的，但也可能有例外。总之，短期稳定性研究应包括一段时间（至少是CRM运输需要用的时间）内运输中可能出现的温度（例如：上至70℃、下至-50℃）。如果这段时间被限制，例如三周，那么短期稳定性研究进行三至四周已足够。

8.3 结果评估

8.3.1 趋势分析

评估稳定性研究数据的第一步是检查数据中是否有任何可观察到的趋势。对于微小的不稳定性问题，由于其内在动力学机理未知，因此，线性拟合是一个合适的模型。而当不稳定性是由一个很好定义的机理引起时，应优先选择该模型而不是（经验的）线性模型，这种模型的数学处理要比直线复杂一些，但也可以用相同的方式评估，即采用F检验来检验趋势的显著性。

在没有严格确定的动力学机理的情况下，稳定性研究的基本模型可表示为：

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \dots\dots\dots (7)$$

式中：

β_0 和 β_1 ——回归系数；

ε ——随机误差分量。

示例：放射性同位素 RM 是具有严格确定动力学机理参数的一个例子，在这种情况下发生放射性衰变。

随机误差分量 ε 可能仅由随机误差组成，也可能包含一个或多个系统因素。在进行稳定性研究时，X为时间，Y为候选CRM的特性值。对于一个稳定的标准物质/标准样品， β_1 的期望值为0。估计参数 β_0 和 β_1 以及计算不同种类方差的表示式的推导，按照推导方差分析表示式相同的途径进行，如参考文献[20]所示。

设有一组 n 对Y对X的观测值，对于每个 Y_i 都有下式：

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \dots\dots\dots (8)$$

由于重复测量、每个时间点使用一瓶以上的样品等，因此对应每个 X_i 经常会得到一个以上 Y_i 值，在特定稳定性研究的模型中应包括这些因素。不过在进行趋势分析时，可以使用各瓶在时间 X_i 时得到的结果平均值。根据本章和参考文献[20]，这些推导可以直接进行。

回归参数可由下列公式得出。

斜率的估计值可按下式计算：

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \dots\dots\dots (9)$$

截距的估计值按下式计算：

$$b_0 = \bar{Y} - b_1 \bar{X} \dots\dots\dots (10)$$

通过误差分析可以计算 b_1 和 b_0 的标准偏差。按下式估计 b_1 的标准偏差：

$$s(b_1) = \frac{s}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \dots\dots\dots (11)$$

式中：

$$s^2 = \frac{\sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2}{n - 2} \dots\dots\dots (12)$$

按下式估计 b_0 的方差：

$$V(b_0) = V(\bar{Y} - b_1 \bar{X}) = s^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] = \frac{s^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} \dots\dots\dots (13)$$

这里要注意： b_1 和 \bar{Y} 是不相关的^[24]。

基于 b_1 的标准偏差，可以作出判断。用公式 (11) 和合适的 t —因子（相关的自由度等于 $n-2$ ），可检验 b_1 的显著性。尽管该方法并不复杂，但是需要计算参数 $s(b_1)$ ，而该参数不能用通用软件计算。不过大部分软件可计算 F —表，该表也能用来评估回归的显著性（见表1）。

表1 线性回归的方差分析表

变差的来源	自由度	平方和 SS	均方 MS	F
回归	1	$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	MS_{reg}	$F = \frac{MS_{reg}}{s^2}$
回归 (残差)	n-2	$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$s^2 = \frac{SS}{n-2}$	
总和, 平均值 \bar{Y} 的修正值	n-1	$\sum_{i=1}^n (Y_i - \bar{Y})^2$		

通常用 $SS(b_1 | b_0)$ 来表示由于回归导致的均方, 读作“扣除 b_0 后 b_1 的平方和”。

回归的均方 (s^2) 是用 σ_{YX}^2 表示的特性的估计值, 称为对于回归的方差 (s^2)。

采用 F 一表可检验比率 $MS_{reg} : s^2$ 的显著性。表1提供了关于自由度的必要信息。

用 F 一表替代 t 一检验法有以下两个优点:

- 大多数软件系统可制出 F 一表;
- F 一表易于扩展到其它回归模式, 这使其适用范围更宽。

不管使用何种检验, 都应注意只有测量的重复性标准偏差 (可能连同瓶间均匀性一起) 足够小时, 结论才有意义。可以证明当重复性标准偏差与材料均匀性研究以及测定 (例如: 测量特性值) 得到的标准偏差相当时, 测量重复性可以满足要求。增加每个时间点测量的瓶数可降低瓶间不均匀性的影响。当 s_{bb} 等于或大于测量重复性时, 这样的处理可能是必须的。如果能观测到趋势, 通常这意味着材料不能被定值。作决定的准则应以 (期望的) 标准物质/标准样品特性值的不确定度、(要求的) 有效期和有效期内的趋势为基础。如果相对于标准物质/标准样品特性值的不确定度而言, 在 (要求的) 有效期内有明显的趋势, 则或者材料由于缺乏稳定性不能被定值, 或者有效期应该缩短。

8.3.2 无趋势时不确定度的评估

稳定性研究包含下列不确定分量:

- 测量重复性;
- 材料不稳定性;
- 测量系统不稳定性 (经典设计中);
- 再现性方面 (如操作者、设备), 包括校准 (经典设计中);
- 瓶间均匀性 (在批定值中)。

从以上列出的内容可以看出，由于同步设计减少了要考察的分量数目，因此，只要可能应首选同步设计而非经典设计。在典型的同步稳定性研究中，只有三个不确定分量，它们可以通过完全嵌套的双因素方差分析分开^{[20], [23]}。在该实验中，单次测量的不确定度可以表示如下：

$$u^2(y_{ijk}) = s_{stab}^2 + s_{bb}^2 + s_r^2 \quad \dots\dots\dots (14)$$

式中：

s_{stab} ——不稳定性引起的标准偏差；

s_{bb} ——瓶间标准偏差；

s_r ——重复性标准偏差；

i ——时间点；

j ——不同的瓶；

k ——重复测量次数。

与均匀性研究的情况一样，估计值 s_{stab} 的质量依赖于 s_{bb} （和 s_r ）。因此，瓶间均匀性影响不稳定性估计值的质量，这是不可避免的，因为这是方差分析的特性^{[20], [30]}。结果的处理可通过双因素方差分析完成，这类似于在附录A.2中描述的情况。应该注意（至少在理论上）通过稳定性研究估计 s_{bb} 是可能的^[23]。当稳定性研究中有多个温度时，在参考温度得到的 s_{bb} 估计值通常是最好的，因为假设在这个温度下材料是稳定的。在材料明显不稳定的温度下，材料的变化可能会影响到 s_{bb} 的估计值。

一般可以假定材料的均匀性和稳定性是独立的，但也可能有例外。当瓶间不均匀性相当大时，由于材料的稳定性依赖于材料的组成，可预计瓶与瓶之间材料的稳定性将不同。一些不稳定组分的存在使得整批非常不均匀，也可能是导致这种相关的原因。

经典设计中，重复测量的不确定度的公式：

$$u^2(y_{ijk}) = s_{stab}^2 + s_{lor}^2 + s_{bb}^2 + s_r^2 \quad \dots\dots\dots (15)$$

式中增加了一项，即由于缺乏重复性而引起的方差⁴⁾ s_{lor}^2 ，该项表示测量系统的稳定性。由于在经典稳定性研究中，测量是在（实验室内）再现性条件下进行的，测量系统的稳定性与候选标准物质/标准样品的稳定性不能区分开而导致复杂化。因此，经典设计中不稳定性的不确定度总是要比同步设计的大。

4) “重复性欠缺”意味着在时间点之间除测量重复性之外还存在一个再现性影响。

另一可选的方法是通过斜率为零的回归直线的不确定度估计稳定性的不确定度^[16]，该方法给出材料可能降解的“保守”估计值。

由相应的均方可以确定各自的预算值 (s_{stab} 、 s_{bb} 、 s_r)，在附录A.2中讨论了双因素方差分析。利用完全嵌套的方差分析设计的更详细的资料见(例如)参考文献[20]和[21]。

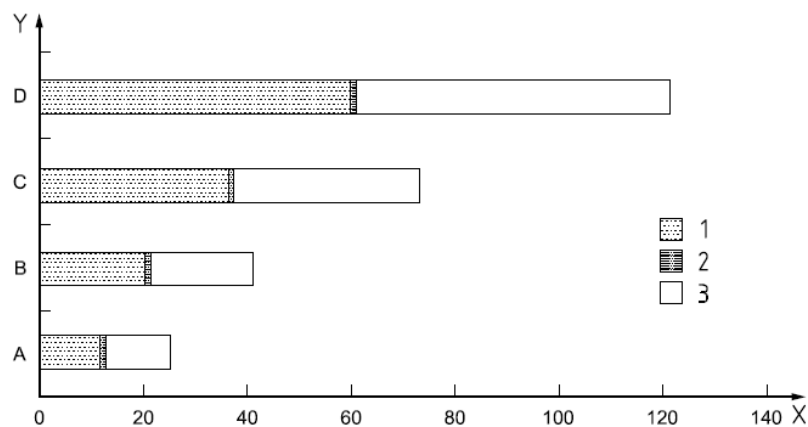
8.4 稳定性监测

8.4.1 实验

在CRM的使用寿命内应重视稳定性监测。在理论上稳定性研究的基本问题考虑的只是过去而不是现在或将来。某些种类的降解或其它不稳定问题进行得非常缓慢、并且是渐进的，但在许多情况下，在某一时间特别是在CRM使用寿命结束时，特性会发生急剧变化。由于这些变化的机理不可预知，因此有必要监测稳定性。

由于同步设计仅在稳定性研究结束时提供数据，而监测必须在CRM的使用寿命内得到有关信息，因此稳定性监测通常采用经典设计。与其它两种稳定性研究不同，由于监测仅涉及证明证书上的不确定度仍然有效，因此，它对CRM特性值的不确定度没有进一步的影响。很显然监测应小心进行，以便在验证CRM的过程中不会引入太大不确定度，但在CRM特性值的合成标准不确定度中不必考虑这些结果。

一种可替代经典监测的重要方法是用同步实验设计进行一种半连续的稳定性研究。图4为这种稳定性监测的一个例子。



说明:

X ——时间, 月

Y ——研究

1 ——准备阶段

2 ——测量

3 ——有效期

图4 半连续稳定性检验

(引自参考文献[23])

在准备阶段，样品储存在相应的温度下。其后，所有样品应在(理想的)重复性条件下测量。有时测量需要一天以上，这意味着要在严格的重复性条件下工作是不可

能的。有效期由该测量数据确定（见8.5）。重要的是要注意，在有效期中止前，应该开展后续研究的测量。由于使用这种设计时估计量只是稍有改进，因此很少需要合并不同稳定性研究得到的结果。

这种稳定性检验不能连续的原因是由于使用了同步测量。由于其是在稳定性检验期后的一次试验中进行，因此应在稳定性检验中设置一个“截止点”。在这个截止点之后CRM特性值的不确定度也可重新评定，因为这些新的稳定性数据可以用作不稳定性不确定度的一个更新的估计值。实际上截止点是由于每个同步稳定性研究都是一次校准完成的，而对于每次新的研究，这种校准是必须的。但在经典设计中，每个数据点通常都要进行一次新的校准，导致“重复性欠缺”。

8.4.2 不确定度评估

稳定性监测的不确定度评估完全不同于长期和短期稳定性研究。首先，应注意稳定性监测不影响标准物质/标准样品证书上不确定度（ u_{CRM} ）的陈述。正如将要证明的一样，在逻辑上这是不可能的，而且也是不必要的。理想情况下，测量不确定度（ u_{meas} ）比 u_{CRM} 小很多，但也不总是这样。此外，测量应以这种方式进行，即其有效性不用CRM来证明。不能用一次实验同时检查这两个方面。只有测量被证明是可靠的，CRM的有效性才能被确认。

如果条件符合：

$$|x_{CRM} - x_{meas}| \leq k \sqrt{u_{CRM}^2 + u_{meas}^2} \quad \dots\dots\dots (16)$$

式中：

x_{CRM} ——CRM的特性值；

x_{meas} ——测量的观测值；

k ——置信水平95%时一个适当的包含因子，那么可以认为材料足够稳定，稳定性得到了证明（假定测量方法是无偏的）。

若这些实验条件全部满足，特性值及其扩展不确定度都将被再次证实。在这些条件下，不需要增加 u_{CRM} ，因为测量的不确定度必须单独考虑。这对于CRM监测和正常使用都是正确的。但应注意，为了监测测量的有效性， u_{meas} 应尽量小，并不应超过典型用户用类似方法验证其测量时的 u_{meas} 值。

如果监测或连续同步稳定性检验显示特性值在其不确定度范围内不再有效的话，有以下两种合理的选择（见6.7）：

- 收回（有证）标准物质/标准样品，或
- 重新定值。

8.5 确定与长期稳定性相关的有效期

有效期与长期稳定性的标准不确定度可能有如下关系，其基础是稳定性数据中无显著趋势。

$$Y(b_0, b', X) = Y_0(1 + b'X) \quad \dots\dots\dots (17)$$

式中假定特性值 Y 作为时间 X 的函数以恒定的相对降解速率 b' 从初始值 Y_0 开始线性减小。CRM 特性值的不确定度可通过变量 Y 所依赖的独立变量 Y_0 、 X 和 b' 的不确定度 $u(Y_0)$ 、 $u(X)$ 和 $u(b')$ 传播来进行估计。

$$u^2(Y) = \left(\frac{\partial Y}{\partial Y_0}\right)^2 \cdot u^2(Y_0) + \left(\frac{\partial Y}{\partial X}\right)^2 \cdot u^2(X) + \left(\frac{\partial Y}{\partial b'}\right)^2 \cdot u^2(b') \quad \dots\dots\dots (18)$$

如参考文献[25]中所示，最后一项可以被认为是长期稳定性引起的标准不确定度与有效期联系起来的基础。应注意该偏微分等于 X （定值后经过的时间），利用一级近似在时间 X 下得到：

$$u_{ls} = Y_0 X u(b') \quad \dots\dots\dots (19)$$

在给定的有效期 X 没有显著降解的情况下，该表示式成为将样品长期稳定性的不确定度考虑进去的依据。

9 特性值测定

9.1 总则

测定特性值有许多技术上有效的途径，包括一个或多个实验室用一种或多种方法进行测量。具体选择何种方式取决于标准物质/标准样品的类型、最终使用要求、参加实验室的资质、所用方法的质量及评估特性值不确定度的实际能力。

本章以及第10章限于测定单个特性值的情况。在有些领域，CRM的特性可以采用其它形式，如光谱。原则上，第9章以及第10章的内容同样也适用于这些种情况，只是数学处理要比特性值的数学处理更复杂。要应用这种概念，需要很好地了解本指南的统计模型。在涉及到光谱或曲线这类特性时，瓶间均匀性、长期稳定性，如果可能，短期稳定性等同样也适用于这种情况。

9.2 建立和证明溯源性

作为测量标准，CRM的特性值要求能够溯源到合适的单位或参考测量标准。要达到这个目的，有几种可能。合适的选择根据CRM的预期用途来定。下面列出几种模式：

- a) 如果可能，CRM 的特性值最好溯源到 SI 单位，并以相应的单位表述。
- b) 许多 RM 是用来（准确）实现由标准方法定义的单位。这些 RM 最好溯源到严格按照标准方法或由标准方法制定的标准操作程序得到的结果。

c) RM 可溯源到其它测量标准或制品, 包括 CRM 和 RM。

也有一些约定标尺至少部分是由标准物质/标准样品保持的, 如pH值及汽油辛烷值的标尺。pH值的测量, 国际上公认的基准复现是用Harned池^[26]。约定标尺通常(但不总)是按照某一固定方法制备的RM保持该标尺。如果可能, 应严格遵循这种固定的方法。

对于许多基体标准物质/标准样品, 情况更复杂。尽管特性值的测量本身可通过如校准所用的仪器而溯源到合适的单位, 但是, 象由一种物理(化学)形态转化为另一种形态的过程却无法溯源。这种转化可以通过与某一测量标准(适用时)比对或在它们自己之间进行比对。对某些转换, 规定了标准方法而且它们可以在定值方案中用以评定与这种转换有关的不确定度。在其它情况下, 只能通过不同实验室用同一种方法进行测量来比对。在这种情况下, 定值是通过独立测量结果之间的一致性来实现的(见第10章)。

测量结果的溯源性通常是通过用适宜的测量标准进行校准来保证的。对许多适用于定值方案的测量系统, 可以通过用测量标准校准仪器来实现。这些测量标准包括另外一些CRM或RM, 这里应注意由于CRM的溯源性有明确说明, 所以最好选择CRM。为了确保仪器校准及结果的溯源性所采取措施的有效性, 可以用如专门研制的控制样品(用于校准以外的样品)进行验证。为此, 这些样品可以没有值(及不确定度)而用于评估校准程序。CRM也可以用于证明测量活动中所得测量结果的有效性。

样品的物理(化学、生物)形态的转化通常是测量方法的重要部分。在某些情况下, 没有合适的方法来验证测量链中的这些步骤。图5总结了一些建立和/或验证测量结果溯源性的典型选择^{[27][28]}。

基体CRM、基体RM及质量控制样品与未知样品并行测量, 可以验证测量结果的有效性。应用这些材料可以支持结果溯源性, 但不能看作是溯源性的直接证明。这些标准物质/标准样品在某种程度上(详细内容见ISO 指南33^[9])控制着测量, 而测量是实现候选标准物质/标准样品定值结果溯源性的先决条件。

空白基体材料及空白提取物等可以用来证明: 当待测特性不存在时, 测量方法所得的结果(象通常成分测量时所做)并非显著不同于零; 或用于校正或建立一个校正因子(+不确定度)。

校准最好用可以溯源到合适参考标准的测量标准来进行。CRM只要满足要求, 就可以用于此目的。校准应适于准确测量, 且不会引入额外不必要的 uncertainty。所选择的参考标准可以是SI单位(如成分分析及许多物理量), 也可以是约定标尺(如方法定义的特性)。

加料样品、加料空白等可以用于确认测量链中的某些部分, 或用于材料的定值过程。建立和证明溯源性这一步达到何种程度取决于这些加料样品是如何制备和赋值的。

为了建立和/或证明测量结果的溯源性，其它需要控制的部分包括：

- 称样
- 试剂、溶剂和“纯材料”的纯度
- 实验室设备及玻璃器皿校准状况
- 测量信号的干扰
- 计算所采用的合理有效的统计/数学技术（如校准曲线、外推法）。
- 测量过程的沾污、损失、缺陷。

以上这些都能通过方法确认进行有效控制。定值方案中使用的任何方法都应是适当确认的，并应可以证明用这种方法所得到的任何结果都能满足该方法在确认期间所建立的技术条件。实验室比对可以是方法确认的一部分。约定方法（即依赖于方法的特性）必须进行实验室间方法确认。其它方法也最好进行实验室间确认。

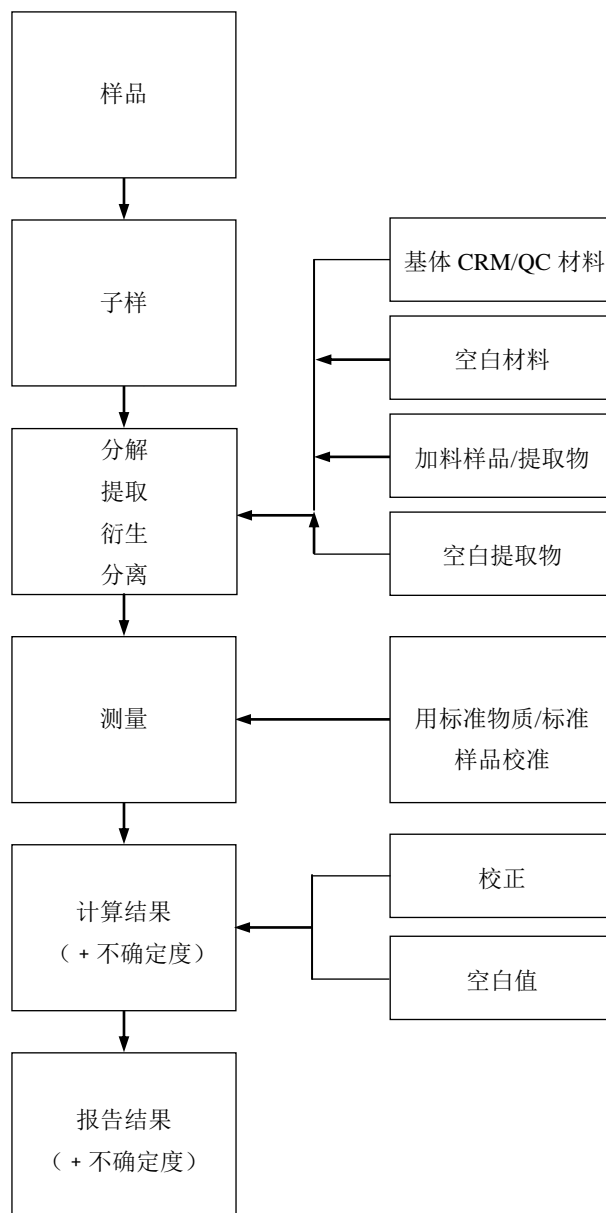


图5 测量链

9.3 实用方法

只要技术上可能，候选CRM通常都是根据准确度确定特性值。因此，标准值通常代表真值的当前最佳估计值。有时，测量值不能被看成“真值”，因而采用由某一特定方法所确定的特性值，这种RM的定值不需要测量活动，只需要陈述给定值及相关测量技术，此时CRM是一种校准物。

标准值对“真值”的偏离不应大于陈述的测量不确定度。特性值所陈述的不确定度应考虑测量过程中系统误差、偶然误差、以及样品间的变异（均匀性）及样品随时间的变异（稳定性）。

GB/T 15000.7—2001^[10]将RM的定值分成以下四种基本方式，生产者及定值单位在实施时会有所不同。

- a) 在单个实验室中用单个（基准）方法测量；
- b) 在单个实验室中用一种或多种独立的标准方法测量；
- c) 由实验室网络用一种或多种可证明其准确度的方法测量；
- d) 由实验室网络，用特定方法测量，只给出该方法评定的特性值。

选择定值方式时，最重要的是要考虑不同的不确定度分量对特性值不确定度的贡献程度。此外，重要的还有赋予的特性值及其不确定度是可以证明的。方式a)通常只限于制品校准。9.5.2给出了对化学成分进行定值的典型的例子，证实方式a)是有效的。对于大多数基体CRM，方式a)有遗漏基体影响和/或干扰的风险，为了减小这种风险，最好有由不同的实验组采用独立（基准）的方法得到至少两组独立的结果。RM测定方法的最佳选择，不仅取决于可用的测量方法，而且也取决于RM基体。

9.4 测量方案设计

9.4.1 一个实验室用一种或多种（标准）方法测量

专门用于方式a)、当然也能用于方式b)的一类重要测量方法是由基准测量方法组成的，在化学测量领域，CCQM将基准方法定义如下^[29]：

“基准测量方法是一种具有最高计量特性的方法，其操作能被完整地描述和理解，可以依据SI单位对其不确定度进行完整陈述。

基准直接法测量未知量不需要用相同量的标准作参照。

基准比率法测量未知量与相同量的标准的比值，它的操作必须可以用一个测量方程式完整描述。”

根据以上定义，基准测量方法^[30]可以作为RM的定值方案之一。但是，并非所有的量都有基准方法可供应用。对于许多RM，待测量的测量非常复杂，无法描述到按SI单位建立完整不确定度陈述所需要的程度。

CCQM确定了几种潜在的基准测量方法^[31]：

- 同位素稀释质谱法（IDMS）；
- 库仑法；

- 称量法；
- 滴定法；
- 凝固点降低法。

注： 预期将来基准方法会增加。

称量法广泛用于溶液及混合气体^[32]的制备，凝固点降低法作为材料纯度（物质的量的分数）^[33]测定的直接方法，IDMS广泛用于材料定值及其它高质量的测量。

9.4.2 多个实验室协作研究

9.4.2.1 概念

以不同实验室间或不同测量方法之间一致性为基础建立起来的测定RM特性值的概念，至少基于两个假设：

- a) 具备一组方法/实验室，其在测定 RM 的特性值方面具有同等的能力，能提供可接受准确度的结果。
- b) 方法（或实验室）内和方法（或实验室）间（隐含的）独立测量结果之间的差异，不管其原因（即测量程序、人员和装备的变化）如何，实质上都是可统计的。

每个方法/实验室的平均值都可看成是材料特性值的无偏估计，而方法/实验室平均值的平均值假设为该特性值的最佳估计值，并且协作研究所得的每个结果均应满足9.2的溯源性要求。对于协作研究的结果分布很不规则的情况，如痕量分析，采用中位值或修剪平均值这种更加稳健的统计量来估计特性值可能更合适。

实际上，适合这样一个计划的方法/实验室的数量是有限的，因此，多数情况下不能完全采用随机设计模型，而且，应当注意假设不同的方法/实验室得到的所有结果属于同一总体要谨慎。即使在当前最好的科学技术条件下，方法性能特性的差异、实验室间不确定度的差异仍然存在，这使得上述假设不成立。

此外，这种方式要有效，必须假设所有方法和实验室得到的所有结果都能溯源到（见9.2）适当规定的测量标准。设计实验室协作研究来获取特性值时，要小心选取这些测量标准。

通常，在这样的协作研究中，需要多个实验室，以使获得数值所必须的操作链中的某些步骤的误差随机化。该步骤一般包括检验用样的转化（例如提取，试样分解），转化后试样的进一步处理（如净化）有时也包括不同测量方法的影响。在这些情况下，由单个方法（和在单个实验室内）所获得结果的测量不确定度无法作全面评估时，本章所描述的方式优于第10章给出的方式。

图6为协作研究测定候选CRM的一般程序，图中的每一步都可以看作是有差别的，都有自己的判断标准，在进行下一步之前必须予以满足。

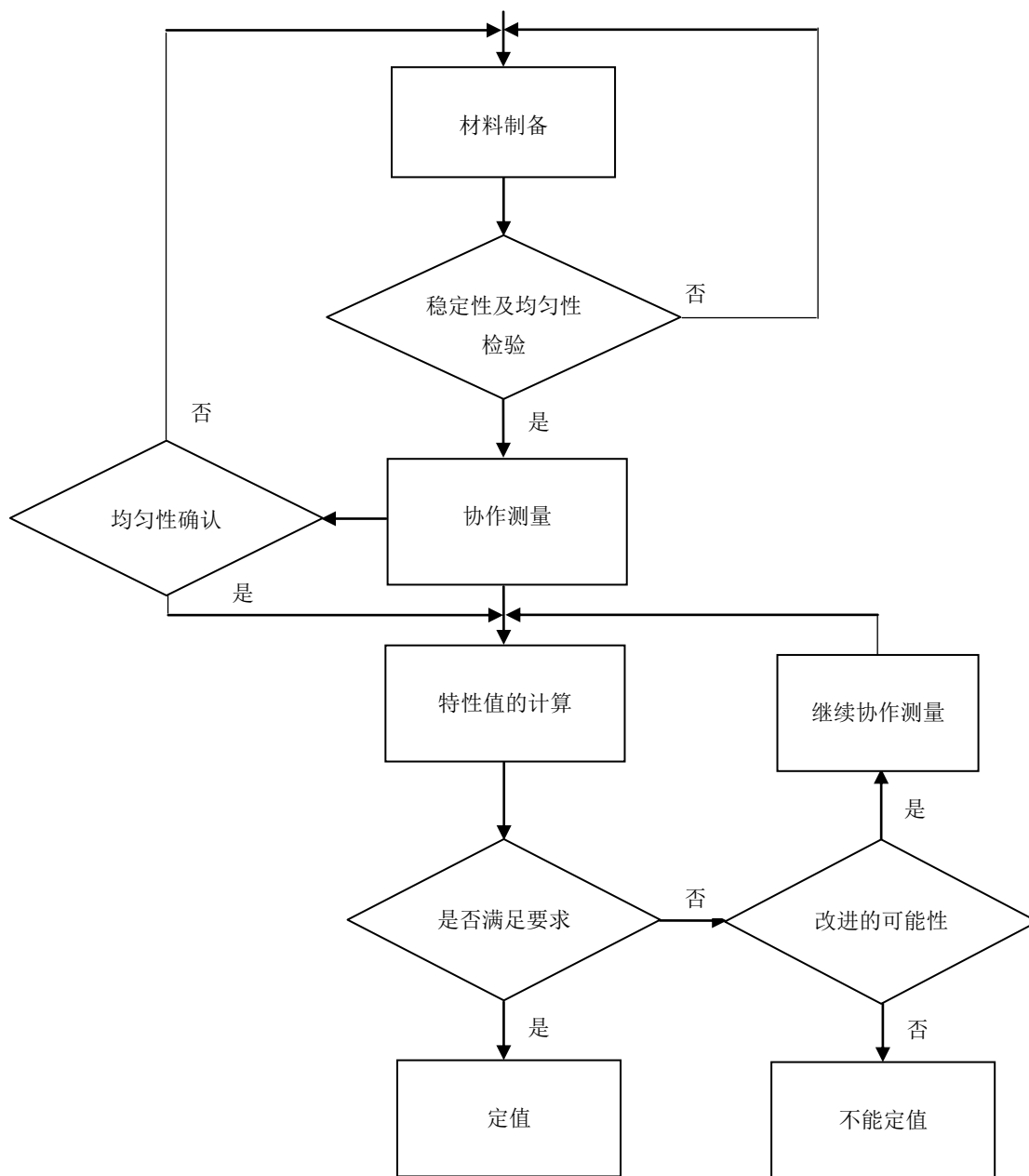


图6 实验室协作分析图

9.4.2.2 管理及进度安排

协作研究管理是负责定值的组织机构的责任。它对所有的相关问题提供足够的指导以确保工作顺利进行。为了获得成功，这种协作研究应有明确的目标、有效的设计和组织、参加实验室和操作人员易于执行的简洁、清晰的指南，无论是操作人员还是实验室，参与这样的项目就意味着同意遵守这些指南。指南应包括时间安排、单元数、每个单元重复测量的次数、测量方法，如果可能，还包括试验用样量等。

T要制定的指南应考虑到9.2、9.3的内容，并把它们转换成明确的指示便于参加

者明白测量结果的质量及其溯源性要求。指导书还应包括验证有关质量、数据所作假设的机制。

组织者应设定时间进度（即发样日期及发样模式），并且要向发样者及收样者提供有关样品如何储存和处理的清晰说明。组织者还应明示报告结果的日期。这个进度应该是在组织者与有关操作者/实验室之间商定的。

9.4.2.3 技术要求

9.4.2.3.1 所需结果数

通常可用来测定某一特性的测量方法非常有限。如果可能，对不同方法所得的结果应进行检查，以检验它们在各自的不确定度范围内是否一致。如果是一致的，就可根据方法之间的这种一致性，算出其平均值。

在协作研究过程中，如果通过把尽可能多的因素随机化来评估RM的测量不确定度，最好选择多个实验室方案。这些实验室也可以是同一单位的不同部门或小组。参与RM测量活动的实验室的最小数目随所需测量程序的复杂程度而变化。测量程序越复杂，实验室间的变差越大。因此，需要增加参加的实验室数目，以得到具有满意的不确定度的特性量值。遗憾的是实际上程序越复杂，有能力完成测定的组/实验室数越少。在某些极端情况下，对某些特定的RM，定值机构不得不完全放弃实验室间的协作。

对于某些特性，如果存在严格确认的方法，参加实验室/组可以少到两至三个。这种情况典型的例子是用基准方法测定基体RM。有时统计和计量上的控制少一些，但（原则上）每个技术上有效的结果仍然适于被采用时，参加实验室的最少数目一般为六至八个。最后，如果有较大的几率得到统计及技术上无效的结果（即有限的统计控制），参加实验室的数目不得少于10个，最好15个。这个最小数目使得数据可以用异常值处理技术进行仔细检查，并使由此确定的特性值能得到合适水平的不确定度。

另外需要考虑的变量是可用方法的数目，以及这些方法在协作研究中的相对表现。在没有基准方法可用时，理想情况是应由六个有能力的实验室/组使用大约三种方法（如可得到）。

9.4.2.3.2 单元数及重复测量的次数

通常，两个单元就已足够，在（至少）两天或两天以上进行六次重复测量。所有的重复测量最好进行独立校准。如果协作研究结果要被用于最终确认RM的均匀性，参加实验室最少应测定三到四个单元的特性量值，以获得足够多的自由度来估算瓶间标准偏差。

9.4.2.3.3 测量方法

当有严格确认的标准测量程序时，协作研究的组织者可以指定参加实验室用单个方法，这也适用于用某一特定方法所定义的特性（例如浸出特性）。否则，组织者应允许参加实验室自己选择方法，只要实验室有证据表明其所选方法的有效性。组织者应从适合某些特性测定的主要方法中找出好的方法，并在有关各方之间就每个实验室

/组将使用什么方法达成一致。

而且，组织者应要求所有用于协作研究的方法都是确认有效的，即协作研究的每个测量结果都能用预定的准则来验证。方法确认研究的一个必要环节是验证其对国际认可标准的溯源、能否溯源到与特定的测量类型相应的水平。在许多新的测量领域，可用的方法有时只经过按ISO 5725^{[1] ~ [6]}的各部分中所述实验室间的研究来确认，这对于方法定义的特性可能已足够。

最后，（在分发样品进行测量前）可召开有关实验室/组会议，便于在协作研究期间进行各项活动，并讨论可能出现的问题或失误。

9.4.2.3.4 报告结果

可以用两种方法来报告结果，这取决于每个实验室是否报出结果的测量不确定度。

如果要求参加组/实验室报出测量结果的不确定度，原则上报出实验结果、扩展不确定度以及包含因子就已足够。但是最好每个实验室都能报告所有不确定度的完整模型，以便评估结果间的任何协方差^{[34], [38]}。

如果不要求不确定度的信息，参加实验室/组应报告独立的测量结果（不是平均值），报告结果的有效数值位数应符合项目指导书。

无论用哪种方法报告结果，最好都要详细报告所用测量程序的摘要，以便于了解测量过程的预处理步骤（如化学分析中样品的分解和待分析成分的分离），同时还要报告相关的参考文献（如适用）。

9.5 有关特性的一些问题

9.5.1 物理特性 RM

通常，对基本单位及其最常用的倍数和分数单位最准确的测量是在国家计量院（NMI）进行，所有误差及不确定度的来源都作了详细考察。测量方法，通常是校准方法，都经过多年的改进，以减少和估计不确定度。这些测量的准确度高，尤其当它们作为比对的对象时。在没有比对时，比如实验室用室内确认方法，测量应该作出限制。无论是何种实验室，与其它实验室比对来证明方法的性能是确保测量结果质量及溯源性的基础，因此任何新建的实验室，都需要广泛的比对以确保它们自己得到的值及其不确定度是等效的，并且所有影响不确定度的重要因素都没有遗漏。

应特别注意不能在校准体系下测量的物理特性。通常来自方法的不确定度评估与定值研究中所得结果的不确定度评定并不是在校准体系下严格建立起来的，这在测定诸如热导率，冲击表面粗糙度、蠕变及压缩强度等特性时应加以考虑。在这些情况下，协作研究的方案（第10章）可能更合适。更复杂的情况是许多测量并没有组织关键比对。在这方面，物理特性的测定和化学成分的测定之间没有根本的区别。

关键比对以及其它形式的实验室比对，提高了单个计量实验室计算不确定度的可信度。理想情况下，应运用这种比对的结果来改进不确定度模型以及变量和/或不确

定度的估计值。即使这些不需要改进，参加这些比对也是衡量一个实验室测量能力的一个重要手段。比对可以发现那些没有适当考虑的误差，以及那些影响测量的参数没有有效地控制和评估的情况。

根据单个（计量）实验室的结果测定标准物质/标准样品的特性，无论如何努力，总是隐含着不可忽视的风险。当对一个物理量及物理特性进行定值时，在主要的实验室之间进行比对，并且和所有参加者对结果进行充分讨论以解决任何可能出现的分歧非常重要。如果NMI没有参加测量，一开始就应建立由参加实验室到有关国家实验室的完全溯源。

如果存在一种以上的方法，而且这些方法同等有效，对它们进行比对是非常重要的。但是，它们应该与其它方法有同等水平的准确度。对候选标准物质/标准样品的定值，最好用更加准确的方法。

极端情况下可能存在这种情况，即一个实验室，已经把他们的方法和所有其它可能的方法做了比对，并消除了大部分误差及误差来源，只要采取众多的措施来避免所有可能的随机误差来源，就能改进它的方法以减小其不确定度。

9.5.2 化学特性 RM

9.5.2.1 纯物质 CRM

在化学上，纯物质是许多溯源链的基础。所谓“纯”是一种理想状态，所有的物质都不是百分百的纯，总是含有少量杂质。在化学测量中，纯度的定值是建立溯源性最重要的基础。CRM或被实验室用于制备校准标准，或被用于其它CRM的定值或制备，如溶液或混合气体。加料（见5.7.4）时必须充分测定所用材料的纯度。而且，为了把质量分数转化为物质的量的分数，必须列出所有的杂质及其质量分数。

除了直接法，如量热法（凝固点降低法）之外，纯度的测定通常用涉及分析化学技术的差减法，具体如下：

- 根据物质的生产制备过程，列出所有可能的杂质清单；
- 测定待测物质中所有可能的杂质；
- 通过差减法计算其纯度。

杂质的测量通常都很难，因为多数杂质的含量都接近检测限。杂质的测量常涉及多个方法及多个实验室，包括9.4.2中所列的方法。相对于杂质的物质质量分数的大小，很可能会引入较高的相对不确定度，而且评估这些不确定度也非常复杂，因为接近数学极限值（物质质量的分数及质量分数只限定在0和1之间）可能产生另外一些问题，包括得到负的估计值的问题（见参考文献[15]附录F）。

主成分的物质质量分数 y 与 k 种杂质的物质质量分数 x_i 函数关系如下：

$$y = 1 - \sum_{i=1}^k x_i \quad \dots\dots\dots (20)$$

假设各杂质物质质量分数之间是独立的（实际情况常如此），主成分的物质质量分数

的合成标准不确定度如下：

$$u^2(y) = \sum_{i=1}^k u^2(x_i) \quad \dots\dots\dots (21)$$

此式是把GUM 的不确定度传播公式应用于该模型而得出的⁵⁾，因而常会发生有些物质质量分数 x_i 为零，原因可能是这种杂质确实不存在，也可能是它们的含量在方法的检测限以下。如果把检出限作为杂质的含量，这个检出限也应作为标准不确定度的一部分，它代表某些不能检出杂质的最高含量。

9.5.2.2 合成 RM 及混合气体

溶液和混合气体等合成标准物质/标准样品广泛用于校准，这些CRM通常采用称量法制备。当采用称量法配制大量溶液时，溶液要经过分样及装瓶等程序，对该批的定值方法如下：

- 第一步：称量组分（这是定值的基础）；
- 第二步：用合适的分析方法验证称量组分；
- 第三步：进行均匀性检验确定瓶间变动性；
- 第四步：进行稳定性检验确定长期稳定性。

第三及第四步对测量不确定度的影响应该较小（见第7、8章），但也应考虑。如可忽略，那么其大小就不会影响CRM特性值的标准不确定度。验证不确定度（一定程度上依赖于成分检验的能力）及称量不确定度一起被计入模型中。CRM特性量值的合成标准不确定度就成为：

$$u_{CRM} = \sqrt{u_{grav}^2 + u_{ver}^2 + u_{bb}^2 + u_{lts}^2} \quad \dots\dots\dots (22)$$

例如，混合气体，已经建立了称量法制备的模型，在某种程度上该模型在此特定领域之外也可应用。GB/T 5274—1985^[35]描述了称量法制备混合气体的配制方法及定值。但对于批量气体标准物质/标准样品，通常用其它方法^[36]。这种批量标准物质/标准样品需要用称量法制备的混合气体作为校准RM来对其进行定值。基于GB/T 5274—1985所描述的方法而确定的详细不确定度模型在参考文献[32]中给出。以制备为基础，混合物的组成可以用母气组分的质量分数来表述，这些母气是用来制备混合气体的气体，它可以是混合气，也可以是纯气（见9.5.2.1）。

GB/T 5274—1985模型考虑了母气中的杂质，这是配制能够溯源到SI单位（即摩尔）的混合气体的重要先决条件。某些影响可能不能考虑进模型中，例如由于吸附和解吸作用，进入气瓶的气体成分和从气瓶中取出的气体成分未必一样。此外，要求对配制过程中所有可能出现的错误进行质量控制核查。为了达到这个目的，常将称量的物质质量分数与测量（验证）的物质质量分数进行比对，并假定配制组分与分析验证的组分没有区别^{[35], [37]}。

5) 这个表达式是一个精确的线性模型，其中 x_i 假设是独立的

9.5.2.3 加料空白及空白基体

用于制备合成CRM的方法同样适用于空白及空白基体的加料。唯一复杂的地方是要验证加料材料是否是真正的空白，或能证明要加入的少量物质是存在的，则这些量应在特性值及其不确定度的模型中加以考虑。

定值方法可能与9.5.2.1相似，但如果试样转化（见9.2）或决定试样转化的不确定度有问题，也可以用9.4中的一种方法。具体选择哪种方法主要取决于9.2和9.3中的考虑因素。

9.5.3 约定特性的测定

在化学、生物化学及其它测量领域，许多特性是由方法试验程序或专用设备决定的，如物质的机械特性及酶的活性等，受沉重的经济负担的影响，这些测量或检测结果可变性很大。

与其它的测量一样，结果取决于程序如何应用。然而，在一些文字标准中，程序的描述并非总是十分详细，操作者无法检验其对程序的理解和应用方式是否正确，因此需要用标准物质/标准样品。

同样，在需要用到某些特定仪器及设备的测量中，可能需要验证仪器是否满足所有的技术条件，但这既昂贵又耗时。要避免这种检验的简单方法就是同时测量或检测已知特性的样品，如果结果满意，就表明仪器处于良好的运行状态，因而可以认为测量结果可以溯源到相关文字标准所建立的测量标尺。

当然，为建立这些特性或测量标尺的CRM，测量需要用到前面所表述的原则。质量、体积、长度或温度等参数的测量必须准确且可溯源，因此需要广泛校准。需要花相当大的力量调查不同程序及测量仪器参数对测量结果的影响。验证及校准即便不在各个实验室，也必须在几个实验室中独立进行，以避免一致性偏倚。当这种偏倚表现为很好一致时，容易被误认为它是正确的。

10 数据及不确定度评估

10.1 模型

即使最具经验的数据评估也会遵循单一或一套规则，但不一定需要用（数学）方程式表示。从技术上考虑，能用数学方程式表达的规则和相互关系被称为模型，两种基本模型论述如下：

- a) 第一种是理论模型，它描述在影响测量程序的变量和待测特性值之间已确立的关系（通常是物理关系）。方程被认为是准确的并用来计算特性值。然后，相同的测量模型被用于传播由有影响的变量赋予的不确定度估计值。
- b) 第二种是经验模型，它描述随机变量和初始（假定）分布特定参数之间假定的相互关系。它们被用来建立程序以确定所涉及的随机变量的合理的统计参数估计值。

还存在上述两种模型的混合形式，通常被称作半经验模型。严格的经验模型可见附录A，其中拟定了一种对不同数量影响因素的方差分析程序。理论和半经验模型以及它们在不不确定度估计中的应用在GUM和参考文献[15]中有描述。

10.2 数据格式

除9.3方式a)的数据格式是由一种方法测得的一个或多个单次测量结果和一个合理的不确定度陈述组成以外（见下面），每个参加实验室报告的数据可以以数据矩阵的形式表示，这个数据矩阵包括：

- 1) 待定特性的适当的估计值（平均值）和不确定度陈述，或者
- 2) 待定特性的一定数量的单次测量结果（重复测量）。

格式1)假定参加实验室具有合适的测量模型能进行基于不确定度的评估（见10.7），而格式2)（更“传统”的一种）需要根据假设的初始分布函数进行统计评估。

由参加实验室提交的结果应根据图7给出的程序评估。

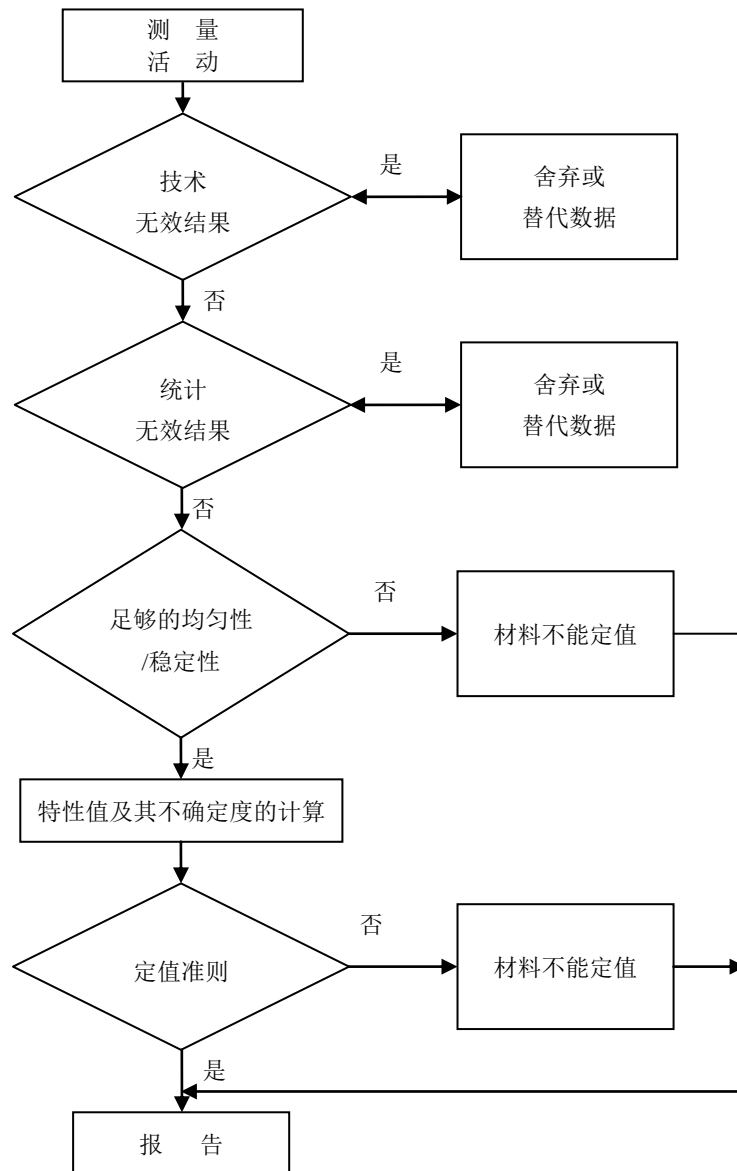


图7 数据处理

为了便于处理及供以后参考, 此类协作研究得到的结果应根据特性分组并系统列表。该表格应包括组/实验室和采用方法的标识、独立的结果、实验室平均值和相应的标准偏差。然而, 如果参加实验室测定了一个以上RM单元的特性值时, 推荐将单元内平均值、总平均值及相应的标准偏差与独立结果分开, 列在另一表格中。当一个参加实验室提交了用不同测量方法得到的多于一组的特性值结果时, 每一组结果都应独立看待, 也就是说应看作如同来自另一个实验室的结果。

也推荐以图表的形式表示结果。

可采用两种报告方法, 这依赖于每个实验室是否报告结果的不确定度。

如果要求参加组/实验室报出它们的测量不确定度, 那么原则上报告测量结果及其扩展不确定度和包含因子即已足够。然而, 最好是每个实验室报告包括所有不确定度的完整不确定度模型。这将方便结果之间任何协方差^{[34]、[38]}的评估。

如果不要求不确定度信息, 参加定值的组/实验室应报告独立结果(而不是平均值)。报告的有效数字位数宜遵守项目规定。

在两种情况下, 建议都要详尽地报告所用测量程序的要点, 以便能了解测量过程的所有前期阶段(例如, 在化学分析中, 样品的分解和待分析物的分离)。(适当时)宜列出参考文献。

10.3 分布

随机变量期望的合适估计量的获得与(假定或确定的)初始分布紧密相连, 该分布大多以概率密度函数表示。在许多情况下, 可以通过绘制直方图, 以图解方法来观察分布。如果不做假设, 只要可以得到统计上合适数量的测量值, 对一定量的测量值的数据评估完全可以建立在:

- 用合适的估计程序(如 Kernel 估计)研究它们的分布, 或
- 根据 GB/T 3358.1—1993(矩计算)中规定的原则计算相关估计量。

其它普遍用于不确定度贡献量化的分布, 例如矩形分布或三角分布也是对称的(见GUM)。

在10.5和10.6中给出的方法总是指数据的正态分布(不包括那些涉及B类不确定度评估的情况)。因此, 为确认所选方法可能需要检验所用数据组的正态性。

如果大部分结果形成两个或多个群, 可以推断方法/实验室间存在不一致, 应考虑以下一些可能性:

- a) 如果这些群与测量方法程序相关, 并且这些群的平均值之间的差异在统计学上和物理学上都是显著的, 那么就没有单一的特性值。在这种情况下, 必需改进测量方法程序来解决这个问题。
- b) 如果这些群与测量方法程序之间不相关, 并且这些群之间的差异在统计学和物理学上是显著的, 就可能需要大量的数据来弥补所用的相对较差的测量方法的不足。

不同方法结果之间的微小差异可以通过引入附加的不确定度分量来解决。这种方法是否有效，只有确定了不确定度分量大小，且验证了由定值设定的标准获得的特性值合成标准不确定度后才能作出判断。解决此类估计问题有几种方法，在参考文献[43]中给出了一些有用的讨论。

如果大部分结果构成了单一的群，可以推断分布是单峰的，很显然待定特性估计量应是结果的平均值。

如果分布是单峰的，应判断正态性假设是否合理，这种判断可以基于直方图的直观观察、正态性检验或者过去关于测定本性的经验。

在某些情况下，结果能被转换以便其服从近似正态分布。一些普遍应用的转换包括对数、平方根和指数的形式。

应注意为了估计特性值及其不确定度，分布函数的假设不及异常值的检验重要。然而，由于大部分描述的统计学是基于正态分布的，因此建议检查数据分布是否符合正态分布，或者转换是否可能，以便可以使用假设正态分布数据的统计技术。

10.4 数据审核

无论何种格式，在应用10.5和10.6程序之前应该用目测和图解法检查数据组。仔细检查观察到的任何异常，找出其是普通的原因（传送错误、打印错误等）还是不寻常的原因（意外丢失、仪器故障等）。如果确认有错误或故障，应剔除相应的结果。

此外，应注意查找技术无效结果。技术无效结果可通过仔细检查测量报告来发现。技术无效结果不一定是统计无效结果。技术无效结果可能很好地落在有效结果的范围内，但是很明显获得结果的条件是不适宜的。技术无效结果应从数据组中剔除。

10.5 数据评估

10.5.1 方式 A: 单个实验室单一方法

当单个值是由附有不确定度陈述的（基准）测量方法得到时，这种具有相应不确定度的值就是测定的结果。如果是一系列的值，通常这一系列值的平均值就是测定的结果。不确定度通常以测量方法的不确定度评估为基础，其值应该与该系列值的标准偏差进行比较，平均值的标准偏差和按测量方法建立的完整不确定度预算应该是一致的。

10.5.2 方式 C: 方法和/或实验室网络

数据格式可以分别是：

- 一系列的值得，每个均附有不确定度陈述；
- 每个实验室的一系列观测值。

在第一种情况下，不确定度陈述作为数据审查的一部分应经过合理性检验，例如通过用测量方法的说明来检验它。然后，数据可按10.7进行处理。

当仅有一系列观测值时，数据可按附录A和参考文献[11]和[12]进行处理。处理程序包括如下步骤：

- a) 检验实验室平均值之间显著性差异以此来决定是否应该求出单个数值的平均值或实验室平均值的均值；
- b) 检验根据 a) 中决定选取的数据组的正态性并找出异常值，异常值可以按 10.5.5 所描述处理；
- c) 检验全部数据组中异常的实验室的方差（如必要，即检出异常方差）。

如果数据组近似正态分布，那么在 a) 中选择的平均值是测定值的选择。如果待测特性值是平均值的均值，即：

$$\bar{Y} = \frac{1}{p} \sum_{i=1}^p Y_i \quad \dots\dots\dots (23)$$

则与平均值的均值有关的合成标准不确定度的主要分量是由下式得到的标准偏差：

$$s^2 = \frac{1}{p-1} \sum_{i=1}^p (Y_i - \bar{Y})^2 \quad \dots\dots\dots (24)$$

而合成标准不确定度为：

$$u_{char} = \frac{s}{\sqrt{p}} \quad \dots\dots\dots (25)$$

在这些公式中，p代表实验室数。这种表达式的有效性大大依赖于Yi值的独立性，所以要验证方法/实验室得到的结果的独立性的假设在何种程度上是有效的，尤其是p数值大的时候，这种验证至关重要（见10.6.2）。

10.5.3 方式 B: 单个实验室的多种方法

也有两种可能的数据格式：

- 一系数值，每个附有不确定度陈述；
- 每个实验室的一系列观测值。

在第一种情况下，对于测定涉及的每种方法来说，都要按10.5.1所述仔细检查其得到的每个值及其不确定度陈述。接着可使用10.7中描述的方法。

10.5.2中的方法可用于第2种情况。至少应可获得方法之一的完整不确定度预算，以便能验证估计的不确定度，并且可以评估测量的偏移。（由于涉及的所有方法均应符合9.2中概述的溯源性要求，因此，这不会导致额外要求。）

10.5.4 方式 D: 方法定义参数

方法定义参数可按10.5.2中所述方法进行处理。

10.5.5 异常数据的处理

如果单个结果或整组结果分别与同组的其它结果或与其它组结果的偏差，不论是精密度还是准确度均大于给定频数分布的统计波动容许值，那么它或它们被怀疑为统计学上的无效结果（异常值）。因此，异常值检验的效力依赖于频数分布假设的有效

性，同样必须要有足够数量的观测值和前后联系的信息以便对异常值作出合适的判断。异常值的检验是统计专家的职责，其不但负责选择有效的异常值检验方法而且负责进行分析。

异常结果可在协作研究的任何阶段发生：单个观测值、成组观测值（例如以瓶为组）或整个方法/实验室得出的结果都可能被观测到是异常的。异常值最好被剔除或者个别情况下（例如计算错误）应该用修正数据代替。只要可能，异常值应根据一种以上异常值检验方法的检验结论才可以剔除。方差的异常值只有在极端的情况下，即当它们与数据群矛盾时才应被剔除。由于获得数据的条件已经与取得所有结果时的条件不同了，因此，补充测量一般是不可接受的。

最后，应注意离散值和异常值的区别（见GB/T 6379.2—2004）^[2]。原则上讲，离散值应在数据组中予以保留，而异常值最好被剔除，以确保在大多数情况下不确定度不被低估。

10.6 不确定度评估

10.6.1 方式 A：单个实验室的单一方法

按10.5.1确定的不确定度为测定不确定度 u_{char} 。

10.6.2 方式 C：方法和/或实验室网络

如果实验室报告的结果附有不确定度陈述可采用10.7中的方法，计算的不确定度是测定不确定度 u_{char} 。

如果实验室仅报告了一系列数值，平均值的标准偏差至少在原理上是测定不确定度 u_{char} ，若不确定度和/或偏移有共同来源，则平均值的标准偏差要加上这些不确定度来源的估计值。

10.6.3 方式 B：单个实验室的多种方法

如果实验室报告的数据附有不确定度陈述，那么可采用10.7中描述的方法，计算的不确定度是测定不确定度 u_{char} 。

在另一种情况下，平均值的标准偏差至少在原理上是测定不确定度 u_{char} 。如果与所用的某种方法得到结果的不确定度进行一致性检查，两者之间存在差异，则在不确定度预算中应考虑到这种差异。

10.6.4 方式 D：方法定义的参数

由方法定义的不确定度评估可用10.6.2所述同样处理。

10.7 基于不确定度的评估

10.7.1 基础

当有不确定度信息时，统计专家的基本任务是把从不同方法/实验室得到的结果及其不确定度合成为单个值（特性值）和合成标准不确定度。当不确定度确定时，可得到作为测量活动一部分的测量不确定度模型。为了获得可能的最佳结果，极力推荐收集这些不确定度模型作为数据评估的一部分。

特性值通常被定义为一些类型的平均值，如有必要，可以使用一些预定的加权模式进行加权处理。当有不确定度模型时，特性值的合成标准不确定度的表达式加上一个附加的不确定度来源（实验室间结果显著分散）后，可以直接使用标准不确定度传播公式获得（GUM:1993，第5章）。这是最简单的方法，但并不总是可以采用。

参考文献[34]中给出了一个可能的替代方法，该方法在参考文献[38]中得到了进一步应用。与特性值相关的合成标准不确定度由下式确定：

$$u_{char} = \sqrt{u_I^2 + u_{II}^2 + u_{III}^2 + u_{IV}^2} \dots\dots\dots (26)$$

式中考虑了4种类型的不确定度：

- 类型 I：一个实验室特有的不确定度；
- 类型 II：所有实验室共有的不确定度；
- 类型 III：实验室各组之间共有的不确定度；
- 类型 IV：参加实验室得到的值之间不一致。

手工进行这种不确定度评估是困难的，并且很容易出错。基于测量模型，只要作为协作研究的一部分收集的信息允许识别模型中的相关的变量，就不必用手工来进行。一旦所有不确定度分量都适当地存储在一个数据库中，这个数据库就可用来导出所有不同类型的不确定度。这可以通过直接确定不确定度的表达式^[34]或者通过 χ^2 拟合⁶⁾ [38], [39]等不同方法来实现。

10.7.2 χ^2 拟合

χ^2 拟合法按如下进行。可以建立如下的矩阵方程：

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_p \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{pmatrix} a \dots\dots\dots (27)$$

式中：

- A ——特性值；
- y_i ——由协作研究获得的测量结果；
- x ——设计矢量。

在这个特别的情况下，只要拟合问题被确定，所有 x_i 值都是已知的。在这种情况下， $x_1 = x_2 = \dots = x_p = 1$ 。结果 y 明显地与不确定度有关。这些将在方差—协方差矩阵中加以考虑，定为：

6) 在文献中，这也被称为“最小二乘”或“最小二乘拟合”。原则上， χ^2 拟合与“最小二乘拟合”差异仅在换算因子上，前者允许拟合程序后进行 χ^2 统计量的评估。

$$V(y) = \begin{pmatrix} u^2(y_1) & u(y_1, y_2) & \dots & u(y_1, y_p) \\ u(y_2, y_1) & u^2(y_2) & \dots & u(y_2, y_p) \\ \dots & \dots & \dots & \dots \\ u(y_p, y_1) & u(y_p, y_2) & \dots & u^2(y_p, y_p) \end{pmatrix} \dots\dots\dots (28)$$

应该注意这个矩阵有 $p \times p$ 阶，并且它是对称的（即上部三角形与下部三角形含有相同的元素）。每两个实验室间的协方差通过研究结果计算中涉及的所有变量和识别共用的变量来评估。在该研究中并不关心这些分量是否只是研究中的两个实验室共用还是更多实验室共用。成对地检查存在的任何协方差的过程意味着在这个过程中，并未区分类型 II 和类型 III 不确定度（见 10.7.1）。

建立了协方差矩阵 $V(y)$ 后，拟合问题可定义如下：

$$\phi(\hat{a}) = (y - X\hat{a})^T V^{-1} (y - X\hat{a}) \dots\dots\dots (29)$$

特性值可示为：

$$\hat{a} = x_{char} = CX^T V^{-1} y \dots\dots\dots (30)$$

式中：

$$C = (X^T V^{-1} X)^{-1} \dots\dots\dots (31)$$

其方差为：

$$V(a) = C = u_{char}^2 \dots\dots\dots (32)$$

有许多固定算法来解决方程（29）中给出的拟合问题。此外，对矩阵 V 作一说明：如果 $V(y)$ 包含与比较相关的所有不确定度分量，则该矩阵等于 $V(y)$ 。否则， $V(y)$ 应加上一个附加的方差—协方差矩阵。这个矩阵的建立类似于 $V(y)$ ，因此通常只有对角线是满的。

这里显示的方法是所述框架的一种简易方法，但这决不是唯一的方法（见参考文献[40]和[41]）。在用该方法之前应先解决异常值的问题（见10.5.5）。同样地，应该严格检查不确定度陈述的可靠性。如有必要，应采用相应加权方案^[42]（见10.5.1）。

众所周知， χ^2 拟合对异常值和加权问题很敏感。

10.8 一些特定问题

10.8.1 应用方差分析法进行数据评估

对于方式C、D和方式B的一部分，方差分析可作为处理数据的工具。在估计瓶间均匀性（例如，见10.8.2和附录A.2）或实验室间标准偏差（见附录A.3）等不确定度分量时，使用方差分析可能特别有帮助。否则，可以采用计算平均值的均值来代替。

10.8.2 应用协作研究进行均匀性确认

协作研究的结果可供最终确认RM均匀性。为此，宜采用双因素嵌套设计，其中应用 pq 个单元数， p 个实验室/方法，各测定 q 个单元的特性值，每个单元重复测定 n 次（见第7章）。重要的是严格遵守这种设计，以满足方差分析的要求和符合其初始的假设^[20]。附录A.2提供了计算相关方差的必要的统计学方法。

10.8.3 特性值的合成标准不确定度

为了确定基于系列测量的特性值，可以采用其它的各种各样的方法。该特性值一般是某种平均值，或有时是稳健平均值^[42]。本节简单地讨论如何将一个平均值的表达式，包括基于方差分析法得到的结果，转换为测定不确定度 u_{char} 表达式。

协作研究的平均值定义如下：

$$\bar{x} = \sum w_i x_i \quad \dots\dots\dots (33)$$

式中：

x_i —— 结果；

w_i —— 权重。

关于权重性质没做任何假设，它们能以观测值的数量、实验室结果的不确定度或其它的方法为依据。然而，这种表达式假设加权方案是以这样一种方式定义的，即它们的总和等于1。

如果所有 x_i 值都有不确定度陈述，并假设所有 x_i 值是相互独立的，那么特性值的不确定度可用下式表示：

$$u^2(\bar{x}) = u_{char}^2 = \sum w_i^2 u^2(x_i) \quad \dots\dots\dots (34)$$

这种基本方法也可用于10.4中给出的方式。当使用不确定度陈述时，应检查它们的可靠性（见10.6.1）。如果得不到不确定度数据，那么平均值的均值的计算可被用于不确定度的估计（见10.5.2和10.6.2）。

11 证书

CRM的概念是作为一类特殊的RM提出的，除了在GB/T 15000.2—1994中定义的RM的特性以外，CRM还附有一个按GB/T 15000.4—2003所述的证书，其中提供了如下信息：

- 目标特性；
- 特性值；
- 特性值的不确定度；
- 特性值计量溯源性的陈述。

市场上有一些RM，它们附有上述CRM信息的文件，但是由于法律或是其它（非技术）原因，该文件不称为“证书”。因为这些RM必须满足作为CRM的相同要求并用于相同的目的，它们也为本指南所涵盖。可以认为“CRM”术语中包含这些RM。本指南

描述制备候选RM的过程以便它们可以被定值,也可以作为带有至少包含如上信息的文件的RM上市销售。

此外,根据GB/T 15000.4—2003, CRM的证书是包含材料选择、适用性确定、待定值特性的测量等广泛详尽的工作程序的概括说明。许多CRM使用者不一定需要比证书更多的信息,但是这些信息可以从定值报告(随CRM提供或承索可得)或向生产者索取得到。证书需有代表定值机构官员的姓名,这表示此人对证书的内容负责。证书是否还应签名最好由定值机构自己决定。

附录 A (资料性附录)

统计方法

A.1 单因素方差分析

假设有 a 个组，每组包含 n_i 个元，理想的情况下各组的元数应该相等，但实际上情况并不总是这样。一些数据可能“缺失”，为计算这些缺失的数据，提出了一些公式^{[20][21]}，用它们来处理不完全的数据组。应当注意，各组缺失的数据越多，估计值的质量就越差。

数据的分散性可用偏差的平方和（简称为“平方和”）表示，这些平方和表示方差分析中不同（等级）水平下的分散性^[20]。由电子表格程序得到的均方可如下转换成方差：

$$s_{within}^2 = MS_{within} \quad \dots\dots\dots (A.1)$$

$$s_A^2 = \frac{MS_{among} - MS_{within}}{n_0} \quad \dots\dots\dots (A.2)$$

式中：

$$n_0 = \frac{1}{a-1} \left[\sum_{i=1}^a n_i - \frac{\sum_{i=1}^a n_i^2}{\sum_{i=1}^a n_i} \right] \quad \dots\dots\dots (A.3)$$

当没有缺失数据时， n_0 即等于 n 。上述原理将测量的分散性归咎于影响材料和测量过程的各个不确定度分量。当组间效应不存在时，预期 s_A^2 是（接近于）零。如果由于实验原因， s_A^2 得到稍负的值时，则将它设定为零。

示例 在瓶间均匀性研究中， s_A 等同于瓶间标准偏差 s_{bb} 。每瓶是一个组。

A.2 数据分析中嵌套的随机效应：双因素方差分析

当测量活动协作研究的结果被用于确认材料的均匀性和确定材料特性值时可以使用这种模式。实验室间研究这一特定情况的试验示意图见图A.1。当测量活动使用不同方法时，其试验计划的设计基本上是一样的。

结果可用下式表示：

$$x_{ijk} = \mu + A_i + B_{ij} + \varepsilon_{ijk} \quad \dots\dots\dots (A.4)$$

式中：

x_{ijk} ——方法/实验室*i*报告的第 *j* 个样品单元的第 *k* 个结果；

A_i ——方法/实验室*i*的误差；

B_{ij} ——方法/实验室*i*内第 *j* 个样品单元的误差；

ε_{ijk} ——测量误差。

要估计的参数是总平均值，实验室间标准偏差 s_L ，瓶间标准偏差 s_{bb} 和重复性标准偏差 s_r 。它们与误差项的关系如下：

$$s_L = \sqrt{\text{Var}(A_i)}$$

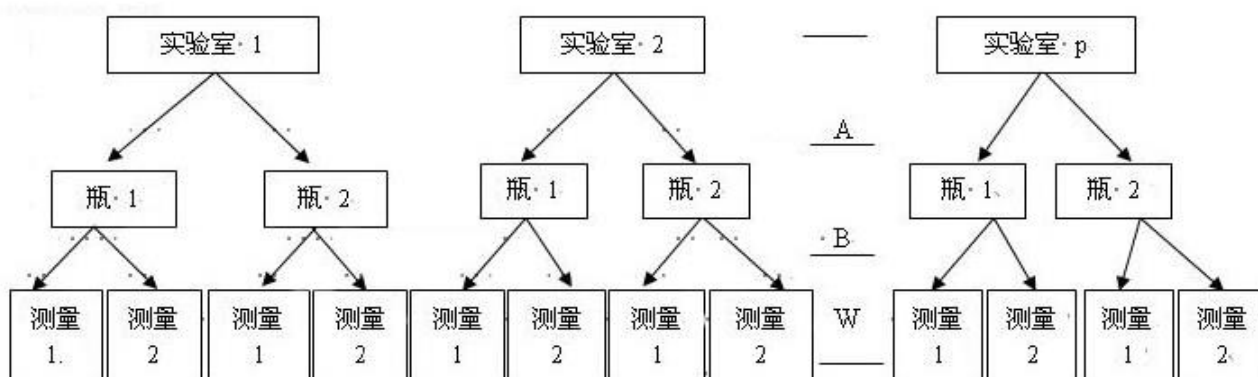
$$s_{bb} = \sqrt{\text{Var}(B_{ij})} \quad \dots\dots\dots (A.5)$$

$$s_r = \sqrt{\text{Var}(\varepsilon_{ijk})}$$

对于瓶间均匀性 s_{bb} ，在均匀性研究不能检出批不均匀性时（见7.9）同样适用。

在剔除技术和统计上的无效结果后，如果有足够数量的相同重复（每个单元相同的重复测定次数和每个方法/实验室相同的单元数）的结果，则所有这些参数可用方差分析法^[20]同时估计。如果由于无效的和/或缺失的结果致使不能满足方差分析的要求，则单元间方差的显著性应采用其它方法确定（见第7章）。

关于均衡和不均衡方差分析的理论和其它方法见标准教科书^{[44], [45]}。RM定值过程中的方差分析讨论见参考文献^{[18], [20], [22], [23]}。



- A 实验室间变差；
- B 瓶间变差；
- W 测量重复性

图 A.1 结合批均匀性研究的协作研究示意图
[RM 的测定（双因素）]

计算上述估计值的公式如下（见参考文献[20]和[21]），总平均值用下式计算：

$$\bar{\bar{x}} = \frac{1}{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij}} \sum_{i=1}^p \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} x_{ijk} \dots\dots\dots (A. 6)$$

式中：

p ——实验室个数；

b_i ——方法/实验室 i 使用的瓶数；

n_{ij} ——瓶 ij 重复测量次数。

方差计算如下：

$$Var(\varepsilon_{ijk}) = MS_{within} \dots\dots\dots (A. 7)$$

$$Var(B_{ij}) = \frac{MS_{B<A} - MS_{within}}{n_0} \dots\dots\dots (A. 8)$$

$$Var(A_i) = \frac{MS_{among} - n_0'Var(B_{ij}) - Var(\varepsilon_{ijk})}{(nb)_0} \dots\dots\dots (A. 9)$$

式中：

$$MS_{among} = \frac{\sum_{i=1}^p n_i (\bar{x}_A - \bar{\bar{x}})^2}{p - 1} \dots\dots\dots (A. 10)$$

$$MS_{B<A} = \frac{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij} (\bar{x}_B - \bar{x}_A)^2}{\sum_{i=1}^p b_i - p} \dots\dots\dots (A. 11)$$

$$MS_{within} = \frac{\sum_{i=1}^p \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_B)^2}{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij} - \sum_{i=1}^p b_i} \dots\dots\dots (A. 12)$$

和

$$n'_0 = \frac{\sum_{i=1}^p \left(\frac{\sum_{j=1}^{b_i} n_{ij}^2}{\sum_{j=1}^{b_i} n_{ij}} \right) - \sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij}} \dots\dots\dots (A. 13)$$

$$n_0 = \frac{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij} - \sum_{i=1}^p \left(\frac{\sum_{j=1}^{b_i} n_{ij}^2}{\sum_{j=1}^{b_i} n_{ij}} \right)}{\sum_{i=1}^p b_i - p} \dots\dots\dots (A. 14)$$

$$(nb)_0 = \frac{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij} - \frac{\sum_{i=1}^p \left(\sum_{j=1}^{b_i} n_{ij} \right)^2}{\sum_{i=1}^p \sum_{j=1}^{b_i} n_{ij}}}{p - 1} \dots\dots\dots (A. 15)$$

均方 (MS) 也可由通用电子表格程序或统计软件得到。所给的表示式用于计算缺失和/或除去 (无效) 数据。对于完整的数据组, 可以使用 ISO 5725-3^[3] 较简单的公式。

A.3 数据分析中嵌套的随机效应: 单因素方差分析

当瓶间均匀性已被其它方法验证时, 采用这种模式 (见第7章)。实验示意图如图A.2所示。此时结果可以简化为:

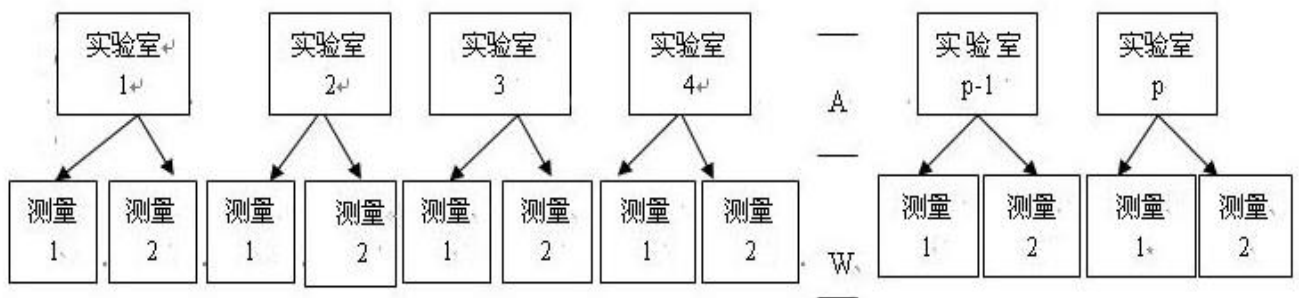
$$x_{ij} = \mu + A_i + \varepsilon_{ij} \dots\dots\dots (A. 16)$$

式中:

x_{ij} ——方法/实验室 i 的第 j 个结果;

A_i ——方法/实验室 i 的误差;

ε_{ij} ——测量误差。



A 实验室间变差;
W 测量重复性

图 A.2 单因素方差分析: 测量协作研究的设计
[RM 测定 (单因素)]

符号的含义与双因素设计的情况一样 (见附录A.2)。要估计的参数为总平均值、

实验室间标准偏差 s_L 和重复性标准偏差 s_r ，它们与误差项的关系如附录A.2中所述。

这些参数的估计值可以用下列公式计算得到^{[20], [21]}，总平均值由下式计算：

$$\bar{\bar{x}} = \frac{1}{\sum_{i=1}^p n_i} \sum_{i=1}^p \sum_{j=1}^{n_i} x_{ij} \quad \dots\dots\dots (A. 17)$$

$$s_r^2 = MS_{within} \quad \dots\dots\dots (A. 18)$$

$$s_L^2 = \frac{MS_{among} - MS_{within}}{n_0} \quad \dots\dots\dots (A. 19)$$

式中符号的含义和附录A.1中的一样。

附录 B（资料性附录）

实例

B.1 总则

本附录的例子主要是想要说明如何使用第7章到第10章的统计方法。因此，大多数的例子都抛开其实际内容，只提供了实践中得到的结果并指出怎样按本指南来处理这些数据。本指南的统计方法本质上是基础性的，它在解决某种特定问题时，并不一定是最合适。在评估数据时还有许多其它问题（例如数据确认、异常值检验、评价不确定度陈述的可信度和软件算法的数字稳定性），这些问题超出了本指南的范围，但它们可能与选择“正确”统计方法同样重要。

有些标准物质/标准样品是利用物理校准或者是利用制备这些标准物质/标准样品的数据来定值的，这种校准的实例在其它公开文献中可以查到，因此本附录中予以省略。在这些项目中遇到的大部分问题主要是与GUM的应用而不是与RM生产的问题有关。

在本指南中给出所有定值项目的实例是不可能的，这些实例只是关系到少数RM生产者。定值报告、公开的文献连同本指南提供的指南构成了定值项目设计的合适基础。

在本附录中对选取的部分定值项目已作了修改以便说明基本的统计方法。适当的例子应考虑到本条款开始所作的说明。应查看例子的合适性，虽然给出了待测特性和基体，但它不会影响数据（例如均匀性研究的数据）评估选择的方法。选择这些例子的唯一依据就是数据大体上呈高斯分布。

B.2 测定模型

CRM IRMM IFCC 452（ γ -谷氨酰转移酶）^[46]的扩展不确定度是通过合成测定、均匀性和稳定性对特性值总不确定度的贡献来估计的：

$$U_{CRM} = k \sqrt{u_{char}^2 + u_{bb}^2 + u_{lts}^2 + u_{sts}^2}$$

包含因子 $k = 2$ 。选定的运输条件无需考虑运输不稳定性造成的附加不确定度，因此 $u_{sts} = 0$ 。

其它不确定度都是相对于测定的特性量值（ x_{char} ）来表示的⁷⁾。该批测定的不确定度 u_{char} 被估计为0.61%。 u_{bb} 和 u_{lts} 的相对估计值分别为0.29%和0.78%。因此，扩展不确定度 U_{CRM} 估计为⁸⁾：

⁷⁾ 只有当所有的不确定度都是相对于同一值表示时，相对不确定度才能被合成。假如有怀疑，最好的作法是合成以绝对单位表示的标准不确定度。

⁸⁾ 该单位 IU/L（国际单位/升）是酶的活性单位。

$$U_{CRM} = 2\sqrt{0.61^2 + 0.29^2 + 0.78^2} = 2.07\% \cong 2.36 \text{ IU/L}$$

该不确定度是根据6年有效期确定的。如果材料的稳定性得到进一步证明，则有效期还可延长^{[46], [47]}。

B.3 均匀性研究

在制备土壤中铬的有证标准物质/标准样品的项目中进行了瓶间均匀性研究，其数据列于表B.1。

表 B.1 土壤中铬的瓶间均匀性研究的测量数据

单位: mg/kg

瓶号	结果 No. 1	结果 No. 2	结果 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

这些数据也能以组平均值、标准偏差和每组的单元数表示。通过计算每瓶平均值、标准偏差和观测次数，由表B.1可以得到表B.2。

表 B.2 每瓶的平均值、方差和测量次数

单位: mg/kg

瓶号	平均值	方差	测量次数
1	123.32	22.54	3
2	120.48	1.20	3
3	122.95	0.26	3
4	118.74	1.10	3
5	111.09	1.20	3
6	119.83	4.95	3
7	118.07	9.81	3
8	122.01	7.04	3
9	116.64	4.18	3
10	125.32	2.45	3
11	124.20	14.85	3
12	124.90	3.45	3
13	125.76	8.76	3
14	120.29	9.50	3
15	118.12	0.47	3
16	120.84	1.06	3
17	120.29	12.54	3
18	125.14	14.06	3
19	131.69	20.08	3
20	122.79	25.76	3

表B.3所示方差分析可以利用诸如电子表格软件计算。

表 B.3 土壤中铬的瓶间均匀性研究的方差分析表

变差源	SS	自由度	MS
瓶间	1037.1	19	54.59
瓶内	330.5	40	8.26
总和	1367.6	59	

瓶间方差用下式计算:

$$s_A^2 = \frac{MS_{among} - MS_{within}}{n_0} = \frac{54.59 - 8.26}{3} = 15.44$$

瓶间标准偏差是该方差的平方根:

$$s_{bb} = \sqrt{15.44} = 3.93 \quad \text{mg/kg}$$

重复性标准偏差可由 MS_{within} 计算:

$$s_r = \sqrt{MS_{within}} = \sqrt{8.26} = 2.87 \quad \text{mg/kg}$$

B.4 应用重复性欠佳的测量方法研究瓶间均匀性

7.9的方法可以举例说明如下。实例取自参考文献[46]。猪肾组织用氯化钠稀释并均匀化。重新沉淀后，产物依次经过DEAE Trisacryl柱和hydroxylapatite柱进行色谱纯化。选择质量浓度为60g/L的牛血清白蛋白作基体，因为它的加入并不改变部分纯化酶的催化活性。通过称量101个每隔一定时间采集的安瓿来检查灌装程序。在灌装期间未发现某种趋势。最后，冻干材料。

检验材料的污染酶、纯度、水含量和残余氧含量，查明材料是适于作CRM的。

灌装程序引起的试样质量变化没有显示不均匀性。为量化单元间变差，评定均匀性的数据取自早先的定值实践。20个安瓿在两天内每天分析三次。汇总2天分析的结果，数据按安瓿分组，并对其进行方差分析。

计算单元内标准偏差 (s_{wb})、单元间标准偏差 (s_{bb}) 以及分析变差对单元间标准偏差的影响 (u'_{bb})。方差分析结果包括不同水平的方差估计值如表B.4所示。

表 B.4 猪肾组织分析结果

平均值	67.78 IU/L
MS_{among}	1.76 IU ² /L ²
MS_{within}	1.63IU ² /L ²
s_r	1.88%
s_{bb}	0.22%
u'_{bb}	0.29%

瓶间标准偏差用下式计算：

$$s_{bb}^2 = \frac{MS_{among} - MS_{within}}{n_0} = \frac{1.76 - 1.63}{6} = 0.0217$$

从而得到 s_{bb} 的估计值为0.147 IU/L，相对值为0.22%。应当注意，每组均包含二个三次重复测量，这就是说 $n = n_0 = 6$ 。重复性标准偏差可由 M_{within} 计算：

$$s_r = \sqrt{MS_{within}} = \sqrt{1.63} = 1.28 \text{ IU/L}$$

应用7.9中的表示式，可以得到瓶间标准偏差的估计值，它反映重复性标准偏差相对较大：

$$u_{bb} = \sqrt{\frac{MS_{within}}{n}} \cdot \sqrt[4]{\frac{2}{v_{MS_{within}}}} = \sqrt{\frac{1.63}{6}} \sqrt{\frac{2}{100}} = 0.196 \text{ IU/L}$$

估计得到的瓶间变差 (s_{bb}) 小于重复性标准偏差 (s_r) 对 s_{bb} 的影响, 这种影响被估计为 0.196 IU/L, 作为瓶间变差的上限。

B.5 稳定性研究和有效期

对土壤中铬的候选标准物质/标准样品进行了稳定性研究。实验数据列于表 B.5。

表 B.5 土壤中铬的稳定性数据

时间 月	铬的含量 mg/kg
0	97.76
12	101.23
24	102.14
36	97.72

由于没有一种物理/化学模型能够真实地描述该候选标准物质/标准样品的降解机理, 故采用直线作为经验模型。事实上, 对于这种基体 (土壤) 中的特性值 (铬含量) 而言, 人们希望截距 (在不确定度内) 等于测定得到的值, 而斜率趋近于零。

斜率可用下式计算:

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{4.74}{720} = 0.006583$$

式中:

$$\bar{Y} = 99.7125 \quad \bar{X} = 18$$

截距由下式计算:

$$b_0 = \bar{Y} - b_1 \bar{X} = 99.7125 - (0.006583 \times 18) = 99.594$$

直线上的点的标准偏差可由下式计算:

$$s^2 = \frac{\sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2}{n - 2} = \frac{15.947}{2} = 7.973$$

取其平方根 $s = 2.8237$ mg/kg, 与斜率相关的不确定度用下式计算:

$$s(b_1) = \frac{s}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} = \frac{2.8237}{\sqrt{720}} = 0.105233$$

自由度为 $n-2$ 和 $p=0.95$ (95%置信水平) 的学生分布 t - 因子等于 4.30。

由于

$$|b_1| < t_{0.95, n-2} \cdot s(b_1)$$

故斜率是不显著的。因而未观测到不稳定性。

另一种可采用的方法是评估最后的回归方差分析表。表B.6列出其结果。

表 B.6 土壤中铬稳定性数据的线性回归方差分析表

	自由度	SS	MS	F	P
回归	1	0.031 205	0.031 205	0.003 914	0.956
残差	2	15.947	7.973 3		
总和	3	15.978			

该 p 值表示回归是不显著的 (对于95%置信水平, $p < 0.05$ 就成为显著的)。

应用8.5节的方法, 有效期 $t = 36$ 个月的长期稳定性的不确定度贡献即成为:

$$u_{ts} = s_b \cdot t = 0.105\ 233 \times 36 = 3.78\ \text{mg/kg}$$

本例的限制因素是相对差的稳定性数据, 可以预期土壤中单纯铬的不稳定性影响是相当小的。

B.6 应用方差分析确定测定值

以下例子取自一种依据IFCC标准方法在37°C下测量 γ -谷氨酰转移酶 (GGT) 的标准物质/标准样品测定。所有实验室都采用同样的标准操作程序 (SOP) 在37°C下测量催化浓度, 定值是基于不同实验室之间结果的一致性。每个参加实验室都收到7小瓶冻干的GGT材料、内部质控样品、SOP、重新制定的程序、结果报告表以及有关要求的信息。在2天内每天处理3小瓶, 并在处理当天测量每瓶的GGT的催化浓度。市场上购买的校准物与该CRM一起分发, 以便有一个独立的质量评定工具。

称量法、容量法和量热法的溯源性已被证明和文件化, 分光光度法的溯源性是用有证的重铬酸钾溶液加以保证。

事先曾决定: 定值测量将不接受实验室间相对标准偏差超过2.5%的结果。有一个实验室超过了此限制, 而该实验室测量质量控制样品的结果也有偏移, 故决定不采用其数据。可接受的十二个实验室得到的独立结果以及实验室六次测量的平均值、标准偏差和相对标准偏差RSD列于表B.7 (结果修约至同样的小数位)。

表 B.7 GGT 的实验室间研究的测定数据

实验室	试验结果 IU/L						平均值 IU/L	SD IU/L	RSD %
	实验室 01	118.1	118.9	119.0	118.1	118.1	119.2	118.6	0.5
实验室 04	112.6	112.6	110.6	114.0	114.0	114.0	113.0	1.3	1.2
实验室 05	111.9	113.7	110.3	112.4	113.0	110.9	112.0	1.3	1.1
实验室 07	111.1	111.4	115.1	109.3	111.0	109.7	111.3	2.1	1.8
实验室 08	113.0	115.0	112.6	112.6	113.7	113.1	113.3	0.9	0.8
实验室 09	113.3	112.4	113.8	110.2	112.5	114.4	112.8	1.5	1.3
实验室 10	114.0	115.3	114.9	113.7	114.3	112.8	114.2	0.9	0.8
实验室 11	116.8	116.9	117.4	116.7	117.0	116.6	116.9	0.3	0.2
实验室 13	112.6	113.0	113.7	111.7	113.6	111.0	112.6	1.1	1.0
实验室 14	114.9	115.5	114.5	115.7	115.5	115.4	115.3	0.5	0.4
实验室 15	117.1	118.6	117.9	116.4	117.7	118.4	117.7	0.8	0.7
实验室 16	113.9	112.5	111.0	111.1	110.8	112.4	112.0	1.2	1.1

单因素方差分析得到的结果列于表 B.8

表 B.8 一由表 B.7 的数据得到的单因素方差分析结果

变差源	SS	自由度	MS
组间	388.64	11	35.33
组内	76.45	60	1.27
总和	465.09	71	

总平均值由下式计算：

$$\bar{\bar{x}} = \frac{1}{\sum_{i=1}^p n_i} \sum_{i=1}^p \sum_{j=1}^n x_{ij} = \frac{1}{12 \times 6} \times 8216.9 = \frac{8216.9}{72} = 114.12 \text{ IU/L}$$

由于数据集是完整的（每组含有 $n=6$ 个结果），总平均值也可以用平均值的平均来计算：

$$\bar{\bar{x}} = \frac{1}{12} \sum_{i=1}^p \bar{x}_i = \frac{1}{12} \times 1369.5 = 114.12 \text{ IU/L}$$

总平均值的不确定度等于

$$u(\bar{x}) = \sqrt{\frac{s_L^2}{p} + \frac{s_r^2}{n \cdot p}} = \sqrt{\frac{5.68}{12} + \frac{1.27}{72}} = \sqrt{0.49} = 0.70 \text{ IU/L}$$

式中:

$$s_r^2 = MS_{within} = 1.27$$

$$s_L^2 = \frac{MS_{among} - MS_{within}}{n_0} = \frac{35.33 - 1.27}{6} = 5.68$$

B.7 应用加权平均确定测定值

要求参加CRM测定的实验室评估其结果的测量不确定度。这些结果的溯源性用下列方法验证：（1）用含有已知量铬的溶液和（2）用充分测定过、用于检查效能如样品的预处理效能的控制样品。应用这些工具，RM生产者首先评价了参加实验室是否能充分控制并掌握其方法。测定的结果汇集于表B.9。第二列 x_i 是结果，第三列 u_i 为与之相关的标准不确定度。根据对控制样品的数据评估，不确定度陈述被假设（和评价）为是可信的。在不确定度陈述不可信时，一般不应根据在此描述的方法对数据进行加权。

表 B.9 土壤中铬的测定数据

单位: mg/kg (干基)

实验室	x_i	u_i	权重 w'_i	最终权重 w_i
1	135	12	0.006 94	0.037 5
2	122	8	0.015 63	0.084 5
3	123	9	0.012 35	0.066 7
4	117	8	0.015 63	0.084 5
5	102	8	0.015 63	0.084 5
6	120	10	0.010 00	0.054 1
7	121	8	0.015 63	0.084 5
8	124	12	0.006 94	0.037 5
9	114	8	0.015 63	0.084 5
10	133	8	0.015 63	0.084 5
11	124	12	0.006 94	0.037 5
12	131	8	0.015 63	0.084 5
13	131	11	0.008 26	0.044 7
14	123	13	0.005 92	0.032 0
15	121	11	0.008 26	0.044 7
16	123	10	0.010 00	0.054 1

w' 列由下式得到:

$$w' = \frac{1}{u^2}$$

为满足权重总和数为单位1的条件，权重 w' 除以它们的和以得到最终的权重：

$$w = \frac{w'}{\sum_{i=1}^p w'_i}$$

这些权重列在表B. 9的最后一列中。平均值由下式计算：

$$\bar{x} = \sum w_i x_i = 121.9 \text{ mg/kg}$$

相关的标准不确定度由下式计算：

$$u(\bar{x}) = u_{char} = \sqrt{\sum w_i^2 u^2(x_i)} = 2.3 \text{ mg/kg}$$

参考文献

- [1] ISO 5725-1, Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions
- [2] ISO 5725-2, Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method
- [3] ISO 5725-3, Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method
- [4] ISO 5725-4, Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method
- [5] ISO 5725-5, Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method
- [6] ISO 5725-6, Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values
- [7] ISO Guide 31, Reference materials — Contents of certificates and labels
- [8] ISO Guide 32, Calibration in analytical chemistry and use of certified reference materials
- [9] ISO Guide 33, Uses of certified reference materials
- [10] ISO Guide 34, General requirements for the competence of reference material producers
- [11] European Commission, “Guidelines for the production and certification of BCR reference materials — Part A: Guide to proposers of reference materials projects”, Doc. BCR/01/97, Brussels (B), 1 September 1997
- [12] European Commission, “Guidelines for the production and certification of BCR reference materials”, Doc. BCR/48/93, Brussels (B), 15 December 1994
- [13] LAMBERTY, A., SCHIMMEL, H., PAUWELS, J., “The study of the stability of reference materials by isochronous measurements”, *Fresenius J. Anal. Chemistry*, 360 (1997), pp. 359–361
- [14] International Laboratory Accreditation Cooperation, “ILAC G12 Guidelines for the Requirements for the Competence of Reference Materials Producers”, ILAC, 2000
- [15] ELLISON, S. L. R., ROSSLEIN, M., WILLIAMS, A. (eds.), “*Eurachem/Citac Guide — Quantifying uncertainty in analytical measurement*”, LGC, London 2000
- [16] GRIEPINK, B., MAIER, E., MUNTAU, H., WELLS, D. E., “The certification of the contents of six chlorobiphenyls (Nos 28, 52, 101, 118, 153 and 180) in dried sludge. CRM 392”, EUR 12823 EN, Luxemburg 1990
- [17] RICE, J. A., “*Mathematical statistics and data analysis*”, 2nd edition, Duxbury Press, Belmont CA, 1995, chapters 3 and 4
- [18] VAN DER VEEN, A. M. H., LINSINGER, T. P. J., LAMBERTY, A., PAUWELS, J., “Uncertainty calculations in the certification of reference materials. 4. Characterization and certification”, *Accreditation and Quality Assurance*, 6 (2001), pp. 290–294

- [19] LINSINGER, T. P. J., PAUWELS, J., VAN DER VEEN, A. M. H., SCHIMMEL, H., LAMBERTY, A., “Homogeneity and Stability of Reference Materials”, *Accreditation and Quality Assurance*, 6 (2001), pp. 20–25
- [20] 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*, 5 (2000), pp. 464–469
- [21] SOKAL, R. R., ROHLF, F. J., “*Biometry*”, 3rd edition, Freeman, New York 1995
- [22] VAN DER VEEN, A. M. H., LINSINGER, T. P. J., PAUWELS, J., “Uncertainty calculations in the certification of reference materials. 2. Homogeneity study”, *Accreditation and Quality Assurance*, 6 (2001), pp. 26–30
- [23] VAN DER VEEN, A. M. H., LINSINGER, T. P. J., LAMBERTY, A., PAUWELS, J., “Uncertainty calculations in the certification of reference materials. 3. Stability study”, *Accreditation and Quality Assurance*, 6 (2001), pp. 257–263
- [24] DRAPER, N. R., SMITH, H., “*Applied regression analysis*”, 2nd edition, J. Wiley and Sons, New York, 1981, chapter 1
- [25] 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”, *Fresenius J. Anal. Chem.*, 370 (2001), pp. 183–188
- [26] BUCK, R. P., RONDINI, S., COVINGTON, A. K., BAUCKE, F. G. K., BRETT, C. M. A., CAMÕES, M. F., MILTON, M. J. T., MUSSINI, T., NAUMANN, R., PRATT, K. W., SPITZER, P., WILSON, G. S., “Measurement of pH. Definition, standards and procedures”, IUPAC Recommendations 2002, *Pure Appl. Chem.*, 74 (2002), pp. 2169–2200
- [27] KING, B., “Metrology and analytical chemistry: bridging the cultural gap”, *Metrologia*, 34 (1997), pp. 41–47
- [28] SCHILLER, S. B., “Statistical aspects of the certification of chemical batch SRMs”, NIST Special Publication 260–125, Gaithersburg MD, July 1996
- [29] Minutes from the fifth meeting (February 1998) of the Consultative Committee on the Quantity of Material (CCQM) of the Bureau International des Poids et Mesures (BIPM), Sèvres, France (1988)
- [30] MAY, W., PARRIS, R., BECK, C., FASSETT, J., GREENBERG, R., GUENTHER, F., KRAMER, G., WISE, S., GILLS, T., COLBERT, J., GETTINGS, R., MACDONALD, B., “*Definition and terms and modes used at NIST for value-assignment of reference materials for chemical measurements*”, NIST Special Publication 260–136, Gaithersburg MD, January 2000
- [31] KAARLS, R., QUINN, T. J., “The Comité Consultatif pour la Quantité de Matière: a brief review of its origin and present activities”, *Metrologia*, 34 (1997), pp. 1–5
- [32] ALINK, A., VAN DER VEEN, A. M. H., “Uncertainty calculations for the preparation of primary gas mixtures. 1. Gravimetry”, *Metrologia*, 37 (2000), pp. 641–650
- [33] BLAINE, R. L., SCHOFF, C. K. (eds.), “*Purity Determinations by Thermal Methods*”, ASTM STP 838, ASTM, Philadelphia, 1984
- [34] PAUWELS, J., LAMBERTY, A., SCHIMMEL, H., “The determination of the uncertainty of reference materials certified by laboratory intercomparison”, *Accreditation and Quality Assurance*, 3 (1998) pp. 180–184
- [35] ISO 6142:2001, Gas analysis — Preparation of calibration gas mixtures —

Gravimetric method

- [36] GUENTHER, F.R., DORKO, W.D., MILLER, W.R., RHODERICK, G.C., “The NIST Traceable Reference Material Program for gas standards”, NIST Special Publication 260-126, Gaithersburg MD, July 1996
- [37] ISO 6143:2001, Gas analysis — Comparison methods for determining and checking the composition of calibration gas mixtures — Comparison methods
- [38] VAN DER VEEN, A.M.H., “Determination of the certified value of a reference material appreciating the uncertainty statements obtained in the collaborative study”, in Ciarlini P., Cox M.G., Filipe E., Pavese F., Richter D., “*Advanced Mathematical and Computational Tools in Metrology, V*”, World Scientific, Singapore, 2000, pp. 326-340
- [39] NIELSEN, L., “*Evaluation of measurement intercomparisons by the method of least squares*”, Technical Report CCEM WGKC 00-13, CCEM Working Group on Key Comparisons, 2000
- [40] COX, M.G., FORBES, A.B., HARRIS, P.M., “*SSfM Best Practice Guide 4 — Discrete modelling*”, Centre for Mathematics and Scientific Computing, NPL, Teddington (UK), Version 1.1, March 2002
- [41] COX, M.G., DAINTON, M.P., HARRIS, P.M., “*SSfM Best Practice Guide 6 — Uncertainty and statistical modelling*”, Centre for Mathematics and Scientific Computing, NPL, Teddington (UK), March 2001
- [42] COX, M.G., “*A discussion of approaches for determining a reference value in the analysis of keycomparison data*”, CISE 42/99, Centre for Mathematics and Scientific Computing, NPL, Teddington (UK), October 1999
- [43] LEVENSON, M.S., BANKS, D.L., EBERHARDT, K.R., GILL, L.M., GUTHRIE, W.F., LIU, H.K., VANGEL, M.G., YEN, J.H., ZHANG, N.F., “An approach to combining results from multiple methods motivated by the ISO GUM”, *J. Res. Natl. Inst. Stand. Technol.*, 105 (2000), pp. 571-579
- [44] JOHN, P.W.M., “*Statistical design and analysis of experiments*”, New York, NY, Macmillan, 1971, p. 76
- [45] SEARLE, S.R., “*Linear models*”, New York, NY, John Wiley, 1971
- [46] LINSINGER, T, KRISTIANSEN, N, SCHIMMEL, H, PAUWELS, J et al. (2000). *Catalytic Concentration of γ -Glutamyltransferase (GGT) determined by the IFCC-Method at 37 °C*. IRMM/IFCC-452, EUR19577EN
- [47] 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), 1986; CEC Report EUR 10628 EN
-