

前 言

本标准等同采用 ISO/IEC 导则 43-1(1997 年版)《利用实验室间比对的能力验证 第 1 部分:能力验证计划的建立和运作》。

GB/T 15483—1999 在总标题《利用实验室间比对的能力验证》下包括以下两个部分:

——GB/T 15483.1—1999 利用实验室间比对的能力验证 第 1 部分:能力验证计划的建立和运作;

——GB/T 15483.2—1999 利用实验室间比对的能力验证 第 2 部分:实验室认可机构对能力验证计划的选择和使用。

本标准是对 GB/T 15483—1995《实验室能力比对检验的开发与运作》的修订,它强调了以能力验证为目的的实验室间比对的操作(其中大部分原则亦适用于其他目的的实验室间比对)。

与 GB/T 15483—1995 相比较,本标准在以下方面作了修改和补充:

——标准的名称由《实验室能力比对检验的开发与运作》改为现名称《利用实验室间比对的能力验证》;

——在形式和内容上,GB/T 15483—1999 分为 GB/T 15483.1 和 GB/T 15483.2 两个部分。与 GB/T 15483—1995 相比,本标准加进了附录 A“处理能力验证数据的统计方法举例”、附录 B“能力验证计划的质量管理”、附录 C“文献目录”三部分内容,同时增加了对要素的描述,使本标准更细化并具有可操作性。

本标准的附录 A、附录 B 和附录 C 都是提示的附录。

本标准和 GB/T 15483.2—1999 自实施之日起取代 GB/T 15483—1995。

本标准由中国实验室国家认可委员会提出。

本标准由中国实验室国家认可委员会和中国标准化与信息分类编码研究所归口。

本标准由中国实验室国家认可委员会和中国标准化与信息分类编码研究所负责起草。

本标准主要起草单位:中国实验室国家认可委员会、中国标准化与信息分类编码研究所、中国计量科学研究院、国家标准物质研究中心、中国石油天然气集团公司。

本标准主要起草人:刘安平、翟培军、茅祖兴、李仁良、王学文、施昌彦、刘智敏、贾莉。

ISO/IEC 前言

ISO(国际标准化组织)和 IEC(国际电工技术委员会)是为了全球标准化而形成的专门体系。作为 ISO 或 IEC 成员的国家机构,通过由从事特定技术领域的各个组织组成的技术委员会参与国际标准的制定。ISO 和 IEC 技术委员会在共同感兴趣的领域中进行合作。同时,其他政府的和非政府的国际组织,通过与 ISO 和 IEC 联系,也参与这项工作。

ISO/IEC 导则 43-1 由 ISO/CASCO Ad HOC(合格评定委员会专家组)编写,是对 ISO/IEC 导则 43 的修订。初稿先在 CASCO(合格评定委员会)成员及 IEC 国家委员会中传阅和征求意见。终稿经 ISO/CASCO 批准,由 IEC 委员会将其作为 ISO/IEC 导则出版。

ISO/IEC 导则 43:1997 的第 1 部分和第 2 部分将取代 ISO/IEC 导则 43:1984。

ISO/IEC 导则 43:1984 是关于制定和实施实验室能力验证的指南,但它对认可机构使用能力验证的结果强调不够。ISO/IEC 导则 43:1997 将在以下三个方面提供指导:

- a) 如何区别能力验证的实验室间比对和用于其他目的的实验室间比对;
- b) 利用实验室间比对的能力验证计划的建立和运作;
- c) 实验室认可机构选择和使用能力验证计划。

ISO/IEC 导则 43:1997 在总标题《利用实验室间比对的能力验证》下,包含下列部分:

- 第 1 部分:能力验证计划的建立和运作;
- 第 2 部分:实验室认可机构对能力验证计划的选择和使用。

ISO/IEC 导则 43:1997 中的附录,对能力验证计划数据的处理提供了统计上的指导,并为编制能力验证计划运作文件(质量手册)提供指导。

引 言

进行实验室间比对有很多目的,可供参加的实验室和其他机构使用。

例如,实验室间比对可用于:

- a) 确定某个实验室进行某些特定检测或测量的能力,以及监控实验室的持续能力;
- b) 识别实验室中的问题并制定相应的补救措施,这些措施可能涉及诸如个别人员的行为或仪器的校准等;
- c) 确定新的检测和测量方法的有效性和可比性,并对这些方法进行相应的监控;
- d) 增加实验室用户的信心;
- e) 识别实验室间的差异;
- f) 确定某种方法的性能特征——通常称为协作试验;
- g) 为标准物质(RMs)赋值,并评估它们在特定检测或测量程序中使用的适用性。

能力验证是为实现目的 a) 而进行的实验室间比对,即确定实验室的检测或测量能力。但能力验证计划的运作也常为上面所列的其他目的提供信息。

参加能力验证计划为实验室提供了一个评估和证明其出具数据可靠性的客观手段。虽然能力验证计划有多种类型(见第 4 章),但大部分都具有对两个或多个实验室的检测和测量结果进行比对的共同特点。

能力验证计划的主要用途之一是评价实验室胜任地进行检测的能力。这可以包括由实验室自身、实验室客户、诸如认可或法定机构等其他机构进行的评估。它是通过实验室检测能力的外部措施来补充实验室的内部质量控制程序的方法。这些活动也补充了由技术专家进行现场实验室评审的技术(现场评审被实验室认可机构经常采用)。就实验室的用户来说,对一个检测或校准实验室能够持续出具可靠结果的信任是非常重要的。为寻求这种保证,用户可以自己对实验室进行评价或采用其他机构的评价结果。

本标准强调以能力验证为目的的实验室间比对的运作,但大部分原则和导则亦适用于其他目的的实验室间比对的运作。

许多实验室认可机构实施它们自己的能力验证计划,也有相当多的实验室认可机构选用由其他机构实施的能力验证计划或其他形式的实验室间比对。GB/T 15483.2 的目的是为实验室认可机构实施能力验证计划时选择合适的实验室间比对提供协调的原则。

大部分评价实验室技术能力的机构要求或希望将能力验证计划中的满意结果作为实验室具备提供可靠结果能力的重要依据(除非能力验证不当)。

然而,需要强调的是,下列两种情况存在着重大差异:

- a) 按照预定的要求,通过评审实验室的整体运作,对实验室能力进行评价;
- b) 实验室参加能力验证的考核结果,可能仅被视为实验室在某一时间、由特定能力验证计划确定的某项(或多项)检测条件下技术能力的信息。

本标准参考了 ILAC(国际实验室认可合作组织)、ISO/TC 69(统计方法技术委员会)、ISO/REM-CO(标准物质委员会)、IUPAC(国际理论化学和应用化学联合会)、AOAC(美国官方分析化学家协会)、ASTM(美国材料试验学会)以及 WECC 和 WELAC(西欧校准合作组织、西欧实验室认可合作组织,现合并为欧洲实验室认可合作组织 EAL)制定的、与能力验证有关的一些指导文件。

中华人民共和国国家标准

利用实验室间比对的能力验证 第1部分:能力验证计划的建立和运作

GB/T 15483.1—1999

Proficiency testing by interlaboratory comparisons—
Part 1: Development and operation of
proficiency testing schemes

代替 GB/T 15483—1995

1 范围

虽然实验室间比对的用法很多,而且在设计和实施方面有所不同,但仍有可能制定组织比对需要考虑的基本原则。本标准规定了这些原则,并且描述了在组织和运作能力验证计划中应考虑的因素。

GB/T 15483.2 描述了评价实验室技术能力的实验室认可机构如何选择和使用能力验证计划。

本标准适用于能力验证的操作者和使用者,诸如参加实验室、认可机构、法定机构以及需要评估实验室技术能力的客户。对于自我评定的实验室尤为有用。但应认识到,能力验证仅仅是有助于在不同检测实验室的使用者之间建立起相互信任的一种机制。

一些认可机构目前的状况是,实验室为其目的定期地参加所接受的能力验证计划。因此,这些计划的操作者在运作一些专业性管理的能力验证计划时,遵守技术要求、统计程序(见附录 A 中的例子)、质量管理(见附录 B 中的指南)的原则是十分必要的。

无法期望不同能力验证组织机构的操作方法完全相同。对于实验室间比对,本标准不给出专门的操作细则。标准的内容仅作为一个根据具体情况可以适当调整的框架,这既适用于参加者少的计划,也适用于参加者多的计划。

本标准不包含某些组织经常采用的一种技术,即通过分发有证标准物质或其他已经很好表征的样品,去评定单一实验室的能力。

附录 C 列出了文献目录。

2 引用标准

下列标准所包含的条文,通过在本标准中引用而构成为本标准的条文。本标准出版时,所示版本均为有效。所有标准都会被修订,使用本标准的各方应探讨使用下列标准最新版本的可能性。

GB/T 15481—1995 校准和检测实验室能力的通用要求

GB/T 15483.2—1999 利用实验室间比对的能力验证 第2部分:实验室认可机构对能力验证计划的选择和使用

JJF 1001—1998 通用计量术语及定义

ISO 3534-1:1993 统计学 词汇和符号 第1部分:概率和一般统计术语

ISO 5725-1:1994 测量方法和测量结果的准确度(真实性和精密性) 第1部分:一般原理和定义

ISO 5725-2:1994 测量方法和测量结果的准确度(真实性和精密性) 第2部分:测定一个标准测量方法的重复性和复现性的基本方法

国家质量技术监督局 1999-03-08 批准

1999-09-01 实施

ISO 5725-4:1994 测量方法和测量结果的准确度(真实性和精密性) 第4部分:测定一个标准测量方法的真实性的基本方法

ISO 9000 质量管理纲要,1994

ISO/IEC 导则 2:1996 标准化及其相关活动的一般术语及定义

《测量中不确定度表示指南》:1993,由 BIPM(国际计量局)、IEC、IFCC、ISO、IUPAC(国际理论化学和应用化学联合会)、IUPAP(国际理论物理和应用物理联合会)、OIML(国际法制计量组织)颁布
注:标准化文摘杂志出版社出版,1995。

(化学)分析实验室能力验证的国际协调议定书 . Journal of AOAC internation, 76, No. 4, 1993. 926—940

基体效应的评价:建议性指南 . NCCLS(临床实验室标准国家委员会)文件 EP-14P, Villanova, PA, 1994

3 定义

本标准采用下列定义。其中一部分采用了有关的国际/国家标准或 ISO/IEC 导则中的定义。

3.1 检测 test

按照规定的程序,由测定给定产品的一种或多种特性、处理或服务组成的技术操作。(ISO/IEC 导则 2:1996)

3.2 检测实验室 testing laboratory

从事检测工作的实验室。(ISO/IEC 导则 2:1996)

注:“检测实验室”一词可在法律意义上或技术意义上应用,或两者兼用。

3.3 检测物品 test item

为进行能力验证而送交参加实验室的材料或制品。

3.4 检测方法 test method

为进行检测而规定的技术程序。(ISO/IEC 导则 2:1996)

3.5 检测结果 test result

严格按照规定的测量方法所得的特性值。(ISO/IEC 导则 2:1996)

3.6 (实验室)能力验证 (laboratory) proficiency testing

利用实验室间比对确定实验室的检测能力。(ISO/IEC 导则 2:1996)

注:在本标准中,“实验室能力验证”一词的含义极为广泛,它包括了诸如:

- a) 定性计划 qualitative schemes——例如要求实验室识别检测物品的某个组分;
- b) 数据转换演练 data transformation exercises——例如提供给实验室多组数据要求进行处理,以获得进一步的信息;
- c) 单件物品检测 single item testing——一件物品按顺序送往若干个实验室,并按时返还组织者;
- d) 单项演练 one-off exercise——就单一事件向实验室发送一个检测物品;
- e) 连续的计划 continuous schemes——按规定的时间间隔,连续地向实验室发送检测物品;
- f) 抽样 sampling——例如要求个人或组织抽取样品,以供进行后续分析。

3.7 实验室间比对 interlaboratory comparisons

按照预先规定的条件,由两个或多个实验室对相同或类似检测物品进行检测的组织、实施和评价。

注:在某些情况下,参加比对的实验室之一可以作为提供检测物品指定值的实验室。

3.8 参考物质/标准物质 reference material (RM)

具有一种或多种足够均匀和很好地确定了特性,用以校准测量装置、评价测量方法或给材料赋值的一种材料或物质。(JJF 1001—1998 中 6.13)

3.9 有证参考物质/有证标准物质 certified reference material (CRM)

附有证书的参考物质,其一种或多种特性值用建立了溯源性的程序规定,使之可溯源到准确复现的

表示该特性值的测量单位,每一种出证的特性值都附有给定置信水平的不确定度。(JJF 1001—1998 中 6.14)

3.10 参考实验室 reference laboratory

为一个检测物品提供参考值的实验室。

注:例如国家校准实验室。

3.11 指定值 assigned value

对于给定目的具有适当不确定度、赋予特定量的值,有时该值是约定采用的。(JJF 1001—1998 中 1.20((量的)约定真值 conventional true value (of a quantity))中的注(1)、注(2))

3.12 溯源性 traceability

通过一条具有规定不确定度的不间断的比较链,使测量结果或测量标准的值能够与规定的参考基准(通常是国家测量标准或国际测量标准)联系起来特性。(JJF 1001—1998 中 6.10)

3.13 协调者 coordinator

负责协调能力验证计划运作中所有活动的组织(或人员)。

3.14 真实性 trueness

由很大一个系列的检测结果得到的平均值与被接受的参考值之间的一致程度。(ISO 3534-1:1993)

3.15 精密性 precision

在预定(规定)条件下所得的独立检测结果之间的一致程度。(ISO 3534-1:1993)

3.16 离群值 outlier

一组值内与其他值不一致的值。(ISO 5725-1:1994)

3.17 极端结果 extreme results

离群值以及与数据中别的量值极不一致的其他值。

注:这些结果对诸如平均值和标准差等总体统计量具有重大影响。

3.18 稳健统计技术 robust statistical techniques

使极端结果对平均值估计值和标准差估计值的影响减至最小的技术。

注:这些技术给极端结果赋予较小的权,而不是将它们从数据组中剔除。

3.19 测量不确定度 uncertainty of measurement

表征合理地赋予被测量之值的分散性,与测量结果相联系的参数。(JJF 1001—1998 中 3.9)

4 能力验证的类型

4.1 总则

能力验证技术根据检测物品的性质、使用的方法和参加实验室的数目而变化。大部分能力验证具有共同的特征,即将一个实验室所得的结果与其他一个或多个实验室所得的结果进行比对。在某些计划中,参加实验室之一可能具有控制、协调或参考的功能。

以下是能力验证计划的一般类型:

4.2 测量比对计划

测量比对计划涉及的被测或被校物品是按顺序从一个参加实验室传送到下一个实验室,这些计划通常具有如下特征:

a) 某检测物品的指定值由某个参考实验室提供,该实验室可能是国家有关测量的最高权威机构。在实施能力验证过程中,特定阶段对检测物品的校核可能是必要的,这样可以确保在整个能力验证过程中指定值无明显变化。

b) 完成按顺序参加的能力验证计划是费时的(有的要若干年),这会造成一些困难。例如:确保物品的稳定性;严格监控物品的传送和各参加者许可的测量时间;在计划实施过程中需将实验室的执行情况向实验室反馈,而不是等到计划结束。此外,以组为基础比较结果可能有困难,因为测量能力彼此接近的

实验室可能相对较少。

c) 各个测量结果要与参考实验室确定的参考值相比较。协调者应考虑各参加实验室声称的测量不确定度。

d) 用于此类能力验证的物品(测量制品)的例子包括参考标准(如电阻器、量规和仪器)。

4.3 实验室间检测计划

实验室间检测计划涉及从材料源中随机抽取次级样品,同时分发给参加检测的实验室共同进行检测。该技术有时也用于实验室间测量比对计划。完成检测后,将结果返回协调机构与指定值比对,以表明各个实验室和整体组的性能。

用于此类能力验证的物品例子包括食品、体液、水、土壤和其他环境物质。在某些情况下,分发的检测物品是以前建立的(有证)标准物质的分离部分。

每轮比对中提供给参加者的整批检测物品必须充分均匀,从而以后识别出的任何极端结果均不能归因于检测物品有显著变异。(见5.6.2和附录A中A4)。

认可机构、法定机构和其他组织在检测领域应用能力验证时,通常采用该类型实验室间检测计划。

一种常用的实验室间检测计划是“分割水平”设计,其中两个分离检测物品具有类似(但不相同)的被测量值水平。该设计用于估算实验室在某个特定的被测量水平下的精密性,它避免了用同一检测物品作重复测量,或在同一轮能力验证中,包含两个完全一致的检测物品所伴随的问题。

4.4 分割样品检测计划

包括一些法定机构在内的用户,经常采用的能力验证的一种特定形式是分割样品检测技术。(勿与4.3中的分割水平计划相混淆。)

典型的分割样品检测计划的比对数据由包含少量实验室的小组(通常只有两个实验室)提供,这些实验室将作为潜在或连续的检测服务的提供者接受评估。

在商业交易中经常采用类似的比对,这时,把表示贸易商品的样品在代表供方的实验室和代表买方的另一个实验室间进行分割。若对供方和买方实验室出具结果出现的显著差异需要进行仲裁,则通常把另一个样品保留在第三方实验室进行检测。

分割样品检测计划包括把某种产品或材料的样品分成两份或几份,每个参加实验室检测每种样品中的一份。与4.3中描述的能力验证类型不同,分割样品检测计划通常只有数量非常有限的实验室参加(经常是两个)。此类计划的用途包含识别不良的精密度、描述一致性偏移和验证纠正措施的有效性。

此计划经常需要保留足够的材料,以便由另外的实验室做进一步分析以解决那些有限数量实验室间发现的差异。

类似的分样检测技术也用于监控临床实验室和环境实验室。比较典型的是,这些计划包含了将某个实验室与另一个或更多其他实验室在宽浓度范围内对若干分样的结果进行比较。在这些计划中,其中的一个实验室由于使用了标准方法和更为先进的设备等,可以被认为是在较高的计量水平(即较低不确定度)上操作,从而在这些比对中,其结果被认为是参考值。对于参与分样数据比对的其他实验室,该实验室可作为顾问实验室或指导实验室。

4.5 定性计划

评价实验室检测能力,并不总是采用实验室间比对(见3.6中注释a))。例如,某些计划为评价实验室表征特定实物的能力而设计的(例如识别石棉的类型、特定病原有机体等)。

这类计划可能包含计划协调者专门制备外加目标组分的检测物品。因此,在性质上这些计划是“定性”的,它们不需要多个实验室的参与或通过实验室间比对来评价一个实验室的检测能力。

4.6 已知值计划

能力验证计划的其他特殊类型,可能包含制备待测的、被测量值已知的检测物品。因而有可能评价某个实验室检测该物品的能力,并提供与指定值比对的数字结果。再次说明,这样的能力验证不需要很多实验室参与。

4.7 部分过程计划

能力验证的一些特殊类型,包含对实验室完成检测或测量全过程中若干部分的能力评价。例如,现有的某些能力验证计划评价的是实验室转换和报告一套给定数据的能力(而不是进行实际上的检测或测量),或根据规范抽取和制备样品或试样的能力。

5 组织和设计

5.1 构架

5.1.1 任何能力验证的设计阶段都要求配备技术专家、统计学专家以及一名计划协调者,以确保计划成功和顺利运作。

5.1.2 通过与这些专家商议,协调者应制定适用于某项具体能力验证的计划。能力验证计划的设计应避免目标含混不清。在计划启动前,其具体方案应取得一致并文件化(见附录B),一般包括下列信息:

- a) 实施验证计划组织的名称和地址;
- b) 协调者以及参与设计和实施验证计划的这些专家的姓名和地址;
- c) 验证计划的性质和目的;
- d) 选择参加者方法的程序,或适当时允许参加所需满足的准则;
- e) 参加计划(部分计划,如抽样、样品处置、均匀性检验和赋值)的实验室名称和地址,以及期望的参加者数量;
- f) 所选检测物品的性质和检测性质,以及是如何考虑做出这些选择的简短说明;
- g) 获取、处置、校核和运送检测物品的方式的说明;
- h) 通知阶段提供给参加者的信息的说明,以及能力验证各阶段日程安排的说明;
- i) 能力验证计划期望的起始日期和目标日期或终止日期,包括参加者进行试验的日期;
- j) 对持续进行的计划,其分发检测物品的频次;
- k) 参加者进行检测或测量可能需要采用的方法或程序的信息(通常是他们的常规程序);
- l) 所用统计分析的概述,包括指定值的确定和离群值的探测技术;
- m) 返回给参加者的数据或信息的说明;
- n) 能力评价技术的依据;
- o) 对检测结果和根据能力验证结果所作结论的公开程度的说明。

5.2 工作人员

5.2.1 参与制定计划的人员在实验室间比对设计、实施和报告等方面应具有足够的资格和经验,或能与具有这种能力的人紧密合作。这些资格和经验应当包括适当的技术技能、统计技能和管理技能。

5.2.2 如5.1.1所述,这些特定的实验室间比对的运作,也需要有对所涉及的方法和程序具有详尽的技术知识和经验的人员予以指导。为此,协调者可能需要列出一个或多个适当人选组成顾问小组,这些人选可以从诸如专业机构、签约实验室(如果有)、计划参加者或数据的最终使用者中选取。

5.2.3 顾问小组的作用可包括:

- a) 制定和评审能力验证计划在策划、执行、分析、报告和效果方面的程序;
- b) 鉴别和评价由其他机构组织的实验室间比对;
- c) 评价与参加实验室能力有关的能力验证结果;
- d) 就能力验证计划所获结果,以及如何将这些结果和实验室评价的其他方面结合运用,向评审实验室技术能力的任何机构提供建议;
- e) 向遇到问题的参加者提供咨询;
- f) 解决协调者和参加者之间的争议。

5.3 数据处理设备

无论使用什么设备,都应能输入所有必要的数据、进行统计分析以及提供及时和有效的结果。校核

数据输入的程序应得到执行,所有的软件都应予验证、支持和备份。数据文件的存储和安全应受控。

5.4 统计设计

5.4.1 所用的统计模式 and 数据分析技术应文件化,并对选用它们的背景材料作简短说明。有关通用统计程序和能力验证数据处理的进一步细节在附录 A 中讨论。

5.4.2 对能力验证计划进行适当的统计设计是至关重要的。应仔细考虑下列事项及其相互影响:

- a) 所涉及的检测的精密性和真实性;
- b) 在要求的置信水平下检出参加实验室之间的最小差异;
- c) 参加实验室的数量;
- d) 待检样品数目和对每一样品进行重复检测或测量的次数;
- e) 估算指定值所使用的程序;
- f) 识别离群值所使用的程序。

5.4.3 在没有 a) 的可靠信息时,某些情况下可能有必要组织一次先导性实验室间比对(协作试验),以获得该信息。

5.5 检测物品的制备

5.5.1 检测物品的制备可以外包,或由协调者承担。制备检测物品的组织应证明其具备该能力。

5.5.2 任何与检测物品有关的、可能影响实验室间比对完好性的条件,诸如均匀性、稳定性、抽样、在转运中可能的损坏及周围环境条件的影响(见 5.6)等均应予以考虑。

5.5.3 计划中分发的检测物品或材料,在性质上通常应与参加实验室的日常检测物品或材料相类似。

注:临床实验室标准国家委员会(NCCLS)在 1994 年于宾西法尼亚州(PA)威兰诺瓦(Villanova)出版的 NCCLS EP-14P 文件中,给出了建立相似性条约的例子。

5.5.4 分发的检测物品数量取决于是否需要覆盖某一组成的范围。

5.5.5 在结果校核完成之前,不应向参加者披露指定值。然而在某些情况下,检测之前告知目标范围也许是适当的。

5.5.6 除了能力验证计划所需要的检测物品外,还可以考虑制备额外数量的检测物品。在评价了参加者所得结果之后,剩余检测物品有可能作为实验室的参考材料、质量控制材料或培训用品。

5.6 检测物品的管理

5.6.1 检测物品的抽取、随机化、运送、接收、识别、标签、储存和处置等程序应文件化。

5.6.2 能力验证中制备散料时,对每一检测参数而言,散料应充分均匀,这样可以使所有的实验室收到被测参数无显著差异的检测物品。协调者应制定用于建立检测物品均匀性的程序文件(见附录 A 中 A4)。只要可能,在检测物品分发给参加实验室之前应作均匀性检验。均匀性程度应达到检测物品间的差异不会对参加者结果的评价产生显著影响。

5.6.3 只要可能,协调者也应提供证明以确保整个能力验证实施过程中,检测物品充分稳定,不会产生任何显著变化。当需要评审不稳定被测量对象时,协调组织可能有必要规定完成检测的日期,以及任何要求的特定预检程序。

5.6.4 协调者应考虑检测物品可能造成的危险,并采取适当措施,告知可能遭受潜在危险风险的任何有关部门(例如检测物品分发者,检测的实验室等)。

5.7 方法/程序的选择

5.7.1 参加者通常能使用他们所选的方法,该方法与其日常使用的程序一致。然而,在某些情况下,协调者可以指示参加者采用特定的方法,这些方法往往是国家或国际上采纳的标准方法,并已通过适当程序(例如协作试验)所确认。

5.7.2 在应用校准程序时,指定值经常是由高等级的校准实验室(往往是国家标准实验室)使用明确并公认的程序,通过测量而得到的参考值。希望参加实验室都采用相同或类似的程序,但这一点对于校准实验室并非总是可行的。

5.7.3 在参加者自由选择所用的方法时,适当情况下,协调者应要求参加者提供他们所用方法的细节以便利用参加者的结果进行比对,并对该方法进行评议。

5.8 能力验证计划的发展

为了确保能力验证计划能适应于技术和科学的发展,需要引入新型样品或新方法、新程序。根据此类计划的结果对各个实验室的能力作早期结论时一定要谨慎。(见 6.4.3)。

6 运作和报告

6.1 协调和文件化

协调者应负责保证计划逐日运作。所有的活动和程序都应文件化,这些文件可以被列入质量手册或由质量手册来补充。(见附录 B)。

6.2 指导书

6.2.1 指导书是提供给参加实验室在计划中须遵循的所有方面的详细指令。例如,这些指导书可以作为某计划条约中的一个主要部分。

6.2.2 应详细阐述可能对所给检测物品或材料的检测产生影响的因素,这些因素包括操作者、物品或材料的性质、设备状态、检测程序的选择和检测的日程。

6.2.3 也可以提供对检测和校准结果的记录和报告的具体指导(例如单位、有效数字的位数、报告格式、结果限期等)。

6.2.4 应告知参加者如同日常检测那样来处理能力验证物品(除非在能力验证设计中有一些可以偏离这些原则的特殊要求)。

6.3 包装和运输

计划协调者应考虑下列有关检测或测量物品分发的一些情况。包装和运输方法必须恰当,并能保护检测物品的稳定性和特性。危险货物的法规或海关要求等可能对运输有些限制。在某些情况下,尤其是在顺序测量比对计划中,实验室自己也要负责运送物品。

全部有关的海关申报单应由协调者完成,以确保将通关中的延误减至最小。该计划应符合国家的和在国际的适用于检测物品运输的法规。

6.4 数据的分析和记录

6.4.1 从参加实验室处获得的结果应予输入和分析,一旦可行即报回实验室。用程序来校核数据输入、传送和随后统计分析的有效性是至关重要的(见 5.3)。建议将数据表、计算机备份文件、打印结果和图件等按规定保存一定时期。

6.4.2 数据分析应产生总计度量值、性能统计量以及与计划的统计模式和目标相一致的关联信息。利用离群值探测试验加以识别,然后剔除,或者最好利用稳健统计量,将极端结果对总计量的影响减至最小。附录 A 中包含有统计评价的概括性建议。

6.4.3 计划协调者应有文件化的准则来处理对能力评价可能不适合的检测结果。例如,就能力验证的目的而言,当检测材料显示出不够充分均匀或稳定时,对被测量不予分级或计分。

6.5 能力验证报告

6.5.1 能力验证计划报告的内容根据具体计划的目的而变化。但应清晰和全面,并包含所有实验室结果分布的数据,以及各参加者能力的说明(见 6.6)。

6.5.2 能力验证计划报告中通常应包含下列信息:

- a) 实施或协调该计划的组织名称和地址;
- b) 参与计划设计和实施的人员姓名和单位(见 5.2);
- c) 报告的发布日期;
- d) 报告的编号和清晰的计划标识;
- e) 所用物品或材料的清晰说明,包括样品制备和均匀性检验的细节;

- f) 参加实验室代码和检测结果;
- g) 统计数据和总览,包括指定值和可接受结果的范围;
- h) 用于确定指定值的程序;
- i) 任何指定值的溯源性和不确定度的细节;
- j) 为其他参加实验室所用的检测方法/程序确定的指定值和总计统计量(若不同的实验室使用不同的方法);
- k) 协调者和技术顾问对实验室能力的评论;
- l) 用于设计和实施计划的程序(可以包括参照的计划议定书);
- m) 用于对数据作统计分析的程序(见附录 A);
- n) 适当时,提出解释统计分析的建议。

6.5.3 对于定期实施的一些计划,简单的报告可能已经足够,因此日常报告中可省去 6.5.2 条建议的许多内容,但在阶段性总结报告中参加者有要求时应将它们包括进去并实施。

6.5.4 在规定的时间内应尽快使报告具有可获性。虽然,按理想状况,所提供的全部原始数据应报给参加者,但在某些庞大的计划中可能做不到。参加者至少应收到以总结形式表达的(例如图表式)所有实验室的结果。在某些计划中,例如长周期的测量比对计划,应将中期报告发送给各个参加者。

6.6 能力评价

6.6.1 需要对能力评价时,协调者应负责确保评价的方法适合于维持该计划的可信性。

6.6.2 协调者可谋求技术顾问的帮助,以对实验室能力的以下方面提供专家评议:

- a) 总体性能与原先期望值(应考虑不确定度)的比较;
- b) 实验室内和实验室间的变异(以及与先前的计划或发表的精密度数据相比较);
- c) 若可行,方法或程序之间的差异;
- d) 误差(指极端结果)的可能来源和改进能力的建议;
- e) 任何其他建议、推荐或一般性评议;
- f) 结论。

6.6.3 在一个特定计划期间或之后,可能有必要定期地向参加者提供各种总结表。对于一个持续进行的计划,这些表可包括各轮中每个实验室能力的最新总结。若有要求,对这样的总结可以做进一步分析并指出其明显趋向。

6.6.4 无论是完成单项计划,还是陆续完成持续计划的各轮之后,都有各种程序用于评审参加者的能力。附录 A 中给出了这种程序的一些例子。

6.6.5 在能力验证中,不提倡对实验室按能力列表排名的方式出具报告。因此,为避免引起误导和造成误解,对排名应当极其慎重。

6.7 与参加者的沟通

6.7.1 应当提供给参加者一套有关参加能力验证计划的详尽信息,例如一份正式的计划议定书。可以通过信件、通讯和/或报告,结合定期的公开会议,与参加者进行后续联络。计划设计或运行中的任何改变都应立即通知参加者。

6.7.2 如果参加者认为对他们在能力验证中的能力评价有误,他们应能向协调者提出。

6.7.3 应鼓励实验室提供反馈,以使参加者为计划的制定作出积极贡献。

6.7.4 与参加者采取的纠正措施相关的程序(尤其与向认可机构反馈有关),在 GB/T 15483.2 中阐述。

7 保密/道德考虑

7.1 记录的保密性

通常,为各个参加者的身份保密是大多数计划的政策。参加者的身份仅为参与协调计划的极少数成

员所知。这一点还应延伸到以后向显示不良能力的实验室提出补救性建议或措施时。在某些情况下,可能要求协调机构向某个特定管理部门报告参加实验室的不良能力,但在参加者同意参加该计划时应告之这种可能性。

一组参加者,出于改进工作而进行讨论和互相帮助的目的,可以选择在组内放弃保密性。

7.2 结果的窜通和伪造

能力验证的目的主要是帮助参加者改善其能力,但在参加者中仍可能有一种倾向,即对其能力提供一个虚假的良好印象。例如,在实验室之间可能发生窜通,以至不提交真正独立的数据。假如实验室日常进行的是单次分析,但在能力验证中要报告的是样品的重复测量的平均值,或就一个特定计划进行的附加的重复试验,这时实验室也可以给出一个其能力的虚假印象。因此,在可行情况下,能力验证应设计为能确保尽可能少地出现窜通和伪造行为。

虽然协调者采取了各种合理的措施以防止欺骗,但值得称道的是参加实验室应有责任避免欺骗行为。

附录 A

(提示的附录)

处理能力验证数据的统计方法举例

能力验证的结果会以多种形式出现,并涵盖了很宽范围的数据类型,是统计分布的基础。用以分析结果的统计技术应适用于每种情况,但由于其种类太多而无法给出规定。

然而,在评价参加实验室的结果时,对于所有的能力验证,有三个步骤是共同的:

- a) 指定值的确定;
 - b) 能力统计量的计算;
 - c) 能力评价;
- 在某些情况下还有:
- d) 检测物品的均匀性和稳定性的初步确定。

本附录给出了所用统计技术的一般准则,籍此按要求指导具体的应用。

由于新的问题、新的形式、人造检测物品、一致性差的方法或多变的实验室程序等原因,与新的实验室间比对计划的一致性最初往往很差。在一致性得到改善之前,协调者也许要使用相对能力(例如百分位)的稳健度量。当实验室间的一致性得到改善并很好地建立了能力验证计划时,统计技术可能还需要改进。

本附录除了考虑用于处理能力验证数据的统计技术之外,并不考虑用于分析研究的统计技术。对于引言中所列的实验室间比对数据的其他用途,或许需要不同的技术。

注:ISO/TC 68目前正在制定一个文件,将对本附录中包含的统计方法提供详细信息。

A1 确定指定值及其不确定度

A1.1 指定值的建立有各种程序,以下按次序列出一些最常用的程序。在大多数情况下,该次序表明指定值的不确定度在逐渐增加。这些程序分别使用下列各值:

- a) 已知值——由专门的检测物品配方(例如用制造或稀释)决定的结果。
- b) 有证参考值——由定义法确定(用于定量检测)。
- c) 参考值——与一个可溯源到一个国家或国际标准的标准物质或标准并行进行分析、测量或比对检测物品所确定的值。
- d) 从专家实验室得到公议值——专家实验室利用已知的具有高精密度和高准确度的、并可与通常使用的方法相比较的有效方法,确定试验中的被测量时,应具有可证明的能力。在某些情况下,这些实验室可以是参考实验室。
- e) 从参加实验室得到公议值——利用 A1.3 中所述的统计量,同时考虑到极端结果的影响。

A1.2 为了公正地评价参加实验室,以及促进实验室间和检测方法间的一致性,应当确定指定值。可以通过选择共同的比对小组。只要可能,还可利用共同的指定值来实现。

A1.3 用公议技术确定指定值时,采用下述统计量可能是合适的:

- a) 定性值——一个预定的多数百分比的公议值(经常表示在标称或次序刻度尺上);
- b) 定量值——对适当比对的“平均”,诸如:
 - i) 可以是加权或变换(例如,修剪平均或几何平均)的平均值;
 - ii) 中位值、众数或其他稳健度量。

A1.4 只要可能,应采用“测量不确定度表示指南”中所述的程序去确定指定值的不确定度。

A1.5 极端结果处理方法如下:

- a) 当参加者的结果被用于确定指定值时,所用的技术应当使极端结果的影响减至最小。这可以用

稳健统计法或在计算之前剔除离群值来实现(见 ISO 5725-2)。在规模较大或日常的计划中,或许可以对离群值进行自动扫描。

b) 若结果作为离群值而被剔除,那么剔除仅仅是为了计算总统计量。在能力验证计划中对这些结果也应当进行评价,并且给出适当的能力比率。

A1.6 其他的考虑如下:

a) 按理想情况,如果用标准公认值或参加者公认值来确定指定值,那么协调者应当有一个程序来确定指定值的真实度和检查数据的分布。

b) 协调者应有按不确定度接受一个指定值的准则。

A2 能力统计量的计算

A2.1 检测单项检测物品的能力

A2.1.1 能力验证的结果经常需要转换为一个能力统计量,以便于说明并与规定的目标进行比对。其任务是以与能力准则作比对的方式,测量与指定值的偏差。使用的技术可以涉及从无处理要求到复杂的统计变换。

A2.1.2 能力度量对计划的参加者应有意义。因此,度量应与检测的应用需要相关并易于理解,或在一个特定的领域内符合传统惯例。

A2.1.3 变动性度量经常用于计算能力统计量及能力验证计划的总结报告中。对一个适当的比对组,变动性度量的常用例子包括:

- a) 标准差(SD);
- b) 变异系数(CV)或相对标准差(RSD);
- c) 百分位,中位绝对偏差或其他稳健度量。

A2.1.4 对定性结果,通常无须计算。

定量结果常用的统计量,按参加者结果变换程度增加的次序列出如下:

- a) 差($x - X$),在此“ x ”是参加者的结果,“ X ”是指定值;
- b) 百分比差

$$\frac{(x - X)}{X} \times 100$$

- c) 百分位或等级;
- d) z 比分数,在此

$$z = \frac{x - X}{s}$$

s 是满足计划要求的变动性的合适估计值/度量。该模式可用于 X 和 s 由参加者结果推导出的情况,或不是从(所有)参加者结果推导出的情况。(例如,规定了指定值和变动性时;参照“(化学)分析实验室能力验证国际协调议定书”中的 4.2。)

- e) E_n 数(典型地用于测量比对计划),

$$E_n = \frac{x - X}{\sqrt{U_{\text{lab}}^2 + U_{\text{ref}}^2}}$$

其中 U_{lab} 是参加者结果的不确定度, U_{ref} 是参考实验室指定值的不确定度。

A2.1.5 考虑如下:

a) 参加实验室的结果与指定值之间的简单差值可能适用于确定能力,并易于参加者理解。量($x - X$)在 ISO 5725-4 中被称为“实验室偏移的估计值”;

- b) 百分比差适用于浓度;
- c) 百分位或等级用于高度分散或偏态分布的结果和次序响应,或用于有限数目的不同响应。该技

术应谨慎使用(见 6.6.5);

d) 根据检测的性质优先或必须使用变换的结果。例如,以稀释为基础的结果是几何刻度尺的一种形式,可用对数进行变换;

e) 如果利用统计准则(例如 z 比分数),那么变动性的估计值应是可靠的,即根据足够多的观察降低极端结果的影响,并达到低的不确定度。

A2.2 综合能力比分数

A2.2.1 在单一能力验证轮回中,根据一个以上的结果就可以评价实验室的能力。这种情况出现在对一个特定的被测量对象,或一组有关的被测量对象有一个以上的检测物品时。这样做就可以对实验室的能力提出一个更为全面的评价。

一些图表方法,诸如尧敦(Youden)图或表示曼德尔(Mandel) h -统计量的图,都是说明能力的有效技术(见 ISO 5725-2)。

例子如下:

- a) 对相同被测量的合成比分数:
- 满意结果的数目;
 - 总计 z 比分数的平均;
 - 平均绝对差(用单位或百分比表示);
 - 总计绝对差(或平方差)。
- b) 对不同被测量的合成比分数:
- 满意结果的数目(或百分比);
 - 平均绝对 z 比分数;
 - 相对于评价限的平均绝对差。

A2.2.2 考虑如下:

- a) 比分数可以变换(如有必要),以便它们全都服从相同的假设分布(例如对 z 比分数的高斯分布或对平方差的 χ^2 分布);
- b) 应对极端值进行检查,它们可能严重影响一个定量的合成比分数。

A3 能力评价

A3.1 初始能力

应在考虑能力度量是否包含某些特点之后,建立评价能力的准则。

A3.1.1 这些特点如下:

a) 专家公议:由顾问组或其他有资格的专家直接确定报告的结果是否与目标相一致。专家公议是评估定性检测结果的典型方式;

b) 与目标的一致性;例如,考虑方法的实施规范和参加者的操作水平;

c) 对比分数的统计确定;准则应适合于每一个比分数。应用比分数的一般性例子有:

i) 对 z 比分数:

$$\begin{aligned} |z| \leq 2 &= \text{满意} \\ 2 < |z| < 3 &= \text{有问题} \\ |z| \geq 3 &= \text{不满意} \end{aligned}$$

ii) 对 E_n 数:

$$\begin{aligned} |E_n| \leq 1 &= \text{满意} \\ |E_n| > 1 &= \text{不满意} \end{aligned}$$

d) 参加者的公议:从一定比率的参加者使用的比分数或结果的范围,或一个参考组得到,诸如:

——满意的中心百分比率(80%,90%或95%);

——满意的单侧百分比率(最低 90%)。

A3.1.2 对于分割样品设计,其任务可能是识别不合适的校准和(或)结果中大的随机波动。在这些情况下,应根据适当数目的结果和一个宽的集中范围来进行评价。图表技术适用于识别和描述这些问题,尤其是按相应的平均值表明实验室之间差异的图表。结果可以把回归分析和残差分析同参数或非参数技术比较。

A3.1.3 只要可能,应采用图表来显示能力(例如直方图、误差条形图、 z 比分数次序图)。这些图表可用于表明:

- a) 参加值的分布;
- b) 多个检测物品的结果间的关系;
- c) 不同分析方法的比对分布。

A3.2 监控整个时间的能力

A3.2.1 一个能力验证计划可以包含监测整个时间的能力的技术。这项统计技术可以使参加者发现其能力的变动性;是否存在整体倾向性或不一致性,以及其能力在什么地方发生了随机性变化。

A3.2.2 图表法能够便于更多读者理解。传统的常规控制图表(Shewhart)很有用,尤其对于自我改进的目的。数据列表和总计统计量提供了更详细的复查。用于评价能力的统计量可用于这些图和表。

A4 检测物品均匀性的初步确定

应使用适当的统计技术评价从检测物品的均匀性检测中得到的数据。在《(化学)分析实验室能力验证国际协调议定书》中的附录 I (检测材料充分均匀的建议程序)阐述了一个合适的方法。

附 录 B

(提示的附录)

能力验证计划的质量管理

应建立和保持一个文件化的质量体系,例如以质量手册的形式。手册概述一些政策和程序,以确保能力验证计划的质量,并获得参加者和参加者中数据使用者的信任。执行计划的组织应依据 GB/T 19000 系列和 GB/T 15481 的适当部分,满足质量管理和技术能力的要求,并获得一个认可机构的认证和/或认可来予以证明。

质量文件应包含下列款项:

- a) 质量方针;
- b) 协调机构的组织;
- c) 人员的培训和资格;
- d) 人员的职责;
- e) 文件控制;
- f) 审核和评审程序;
- g) 能力验证计划的目标、范围、统计设计和格式;
- h) 运行程序;
 - 样品制备;
 - 样品的均匀性检验;
 - 设备;
 - 建立指定值的程序;
 - 供方,包括分包方;
 - 后勤;

- 数据的分析；
- i) 报告的编制和发放；
 - j) 参加者的行动和反馈；
 - k) 记录的文本和存档；
 - l) 申诉的处理程序；
 - m) 有关保密和道德程序的政策；
 - n) 计算的信息；
 - o) 安全和其他环境因素；
 - p) 分包；
 - q) 参加费用；
 - r) 计划可利用的范围；
 - s) 参加及使用计划结果的总政策。

附录 C
(提示的附录)
文献目录

- [1] WECC 文件 15 : 1992. WECC. 实验室间比对. 西欧校准合作组织 (现为 EAL)
- [2] EAL-G6 : 1993. 能力验证认可标准. 欧洲实验室认可合作组织 (EAL)
- [3] APLAC 实验室间比对
文件 1——校准和测量验证程序
文件 2——检验程序
亚洲-太平洋地区实验室认可合作组织, 1996
- [4] ASTM E 1601 : 1995. 制定和实施实验室能力验证程序的标准指南
- [5] ISO 导则 30 : 1992. 与标准物质有关的常用术语和定义
- [6] Horwitz W. 实验室间分析研究的术语. *Pure Appl. Chem.*, 66(9), 1994, 1903~1911
- [7] Sweets J. A. 和 Pickett R. M.. 诊断系统的评价——单个检测理论的方法. Academic Press, New York, London, 1982
- [8] Youden W. J.. 协作试验的统计技术. Association of Official Analytical Chemists, Arlington, VA, 1967
- [9] Mandel J.. 双因素设计的分析. Chapman & Hall, New York, NY, 1996. 65~79
- [10] Tholen D. W.. 实验室间比对计划中测量和评价能力的统计程序. *Arch. Pathol. Lab. Med.*, 112, 1988. 462~470
- [11] Kafadar K.. 对单个样品问题的双加权法. *Amer. Statistical Assoc.*, 77(378), 1982. 416~424
- [12] Royal Society of Chemistry. 稳健统计——如何不舍弃离群值, 第 1 部分: 基本概念; 第 2 部分: 实验室间试验 (皇家化学协会, 分析方法委员会) *Analyst*, 114, 1989. 1693~1702
- [13] Tholen D. W.. 能力验证中的标准值和作为目标的参加者平均值. *Arch. Pathol. Lab. Med.*, 117, 1993. 885~889
- [14] Ross J. W.. 美国病理学家学院调查中评价标准的演变. *Arch. Pathol. Lab. Med.*, 112, 1988. 334~339
- [15] Hoeltge G. A., Duckworth J. K.. 美国病理学家学院认可的实验室能力验证的评述. *Arch. Pathol. Lab. Med.*, 111, 1987. 1011~1014
- [16] Jenny R. W., Jackson K. Y.. 作为常规检测病人茶碱含量准确度预测值的能力验证. *Clin. Chem.*, 39(1), 1993. 76~81