

过程控制

pH 测定检验能力验证及评价新方法的应用

赵新玥^{1,2}, 赵文², 陈华^{2*}

(1. 无锡市药品安全检验检测中心, 无锡 214028; 2. 中国食品药品检定研究院, 北京 1000502)

摘要 目的: 评价参与能力验证实验室的 pH 测定能力, 探讨能力验证结果评价的方法。方法: 根据 CNAS (中国合格评定国家认可委员会) 规定的程序组织能力验证, 借助 JMP 软件对制备样品进行均匀性和稳定性检验, 通过分析专家实验室的重复性和再现性标准差来确定能力评定的标准差, 以 Z 比分数对结果进行评价。结果: 共有 251 家参加实验室, 实验室总的满意率为 86.45%, 可疑率为 4.78%, 不满意率为 8.76%。结论: 在本次能力验证中, 使用了新统计评价方法与统计工具, 与国内外常用方法相比具有显著优势。
关键词: 能力验证; Z 比分数; JMP; 精密度试验; 柯克伦检验; 格拉布斯检验

中图分类号: R 917

文献标识码: A

文章编号: 0254-1793 (2018) 03-0545-06

doi: 10.16155/j.0254-1793.2018.03.26

Application of new methods to the proficiency testing project of pH determination and its evaluation

ZHAO Xin-yue^{1,2}, ZHAO Wen², CHEN Hua^{2*}

(1. Wuxi Institute for Drug Control, Wuxi 214028, China; 2. National Institutes for Food and Drug Control, Beijing 100050, China)

Abstract Objective: To evaluate the capacity of laboratories which participated in the proficiency testing for pH determination and explore the statistical methods for proficiency testing. **Methods:** The study was performed according to the procedure of China National Accreditation Service for Conformity Assessment (CNAS). JMP was used to evaluate the homogeneity and stability of the test samples. All results conformed to the requirements. Z-scores were calculated using the results of expert laboratories, assigned value and the standard deviation for proficiency assessment. These scores were used to evaluate results. **Results:** A total of 251 labs participated. The satisfaction rate of the laboratory was 86.45%, the suspicious rate was 4.78%, the rate of dissatisfaction was 8.76%. **Conclusion:** In the proficiency testing, we used the new statistical evaluation method and statistical tools. The new method has a significant advantage compared with common methods.

Keywords: proficiency testing; Z-score; JMP; precision experiment; Cochran-test; Grubbs-test

能力验证是利用实验室比对, 按照预先制定的
准则评价参加者的能力^[1]。实验室能力验证最早可

追溯至 20 世纪 40 年代^[2-5], 全球第一家能力验证
服务平台是美国病理学会 (CAP), 全世界已有将近

* 通信作者 Tel: (010) 67095458; E-mail: chenhu@nifdc.org.cn

第一作者 Tel: 15201506801; E-mail: 2460053641@qq.com

30 000 家实验室参与 CAP 的能力验证计划^[6-8]。能力验证对参加实验室自身能力建设有重要的作用与意义也是认可机构加入和维持国际相互承认协议(MRA)的必要条件之一^[9]。能力验证是外部质量评价与自我监控的结合,通过能力验证,可以持续提高实验室的检测能力和水平,保障实验室检测质量,扩展实验室影响。

我国一些行业较早就开始了行业内的实验室间比对活动,但是真正符合国际要求实验室能力验证起步较晚^[10]。中国食品药品检定研究院于 2013 年从中国合格评定国家认可委员会获得能力验证提供者的资格^[11],组织实施药品检验检测领域的能力验证工作。近年来,随着对检验检测质量的重视,能力验证活动日益增多,对能力验证中采用的方法也日益多样化。

本次研究通过组织缓冲液 pH 测定的能力验证计划,采用新的统计分析方法与统计工具对结果进行评价。

1 方法

1.1 项目流程

影响 pH 测定的因素有许多,比如环境、电极、校正溶液、人员技能和仪器性能等等,为了集中反映参加实验室的检测能力,本次检验项目中以具有一定缓冲能力的缓冲液作为样本,具体的流程见图 1。

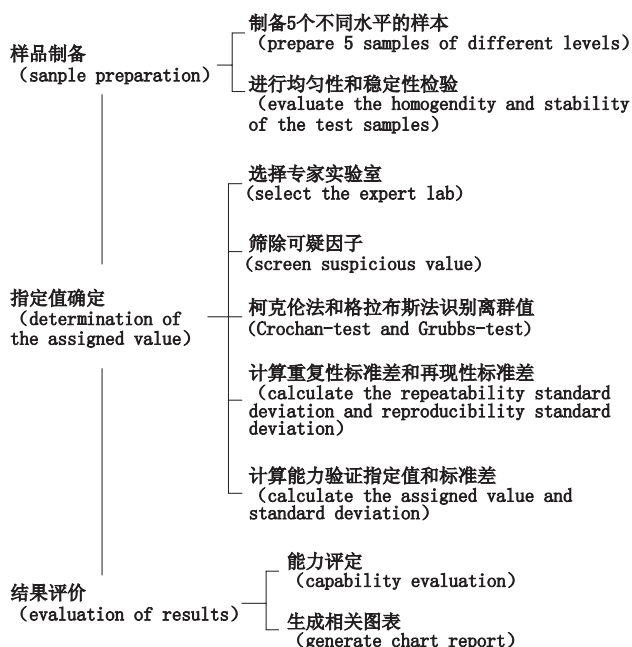


图 1 能力验证流程图

Fig. 1 Flow chart of proficiency testing

1.2 样品制备

本次能力验证用样品均为市售 Reagecon 公司生产的标准缓冲液,从 5 个不同水平样本中给参加实验室选取两个检测水平(缓冲液 1 与缓冲液 2),分别封装于惰性塑料瓶中。

1.3 样品均匀性检验

在所有样品中,以随机方式分别抽取 2 个水平样品 15 瓶,每瓶重复测定 3 次。采用单因子方差分析对检验中的结果进行统计处理。若样品之间无显著性差异,则表明样品是均匀的。判断结果时,根据单因子方差分析法原理,如 $F < F_{\alpha}(f_1, f_2)$ 及显著性水平 $\alpha = 0.05$ 的临界值 F ,则表明样品内和样品间无显著性差异,样品是均匀的^[12]。

1.4 稳定性检验

2 个水平取完整包装的样品各,放置条件分别为高温(60 °C)、高湿(80%RH),分别于 0、1、3、7、14 d 5 个时间点取样,每个时间点取样 3 份,每份测量 3 次。若样品之间无显著性差异,则表明样品是均匀的。判断结果时,根据 t 检验法原理,对每一对结果分别计算 t 值,如果该值小于显著性水平 $\alpha = 0.05$ 的临界值,则表明在测定时间内该样品是稳定的。

1.5 检测方法

参加实验室收到样品后建议参考中国药典 2015 年版四部通则 <0631>(pH 测定法)项目要求实验。考虑到酸度计更新换代,也可参照美国药典 <791> 或欧洲药典 2.2.3 进行仪器校正。测定时,每次单独取样,在重复性条件下,每个水平测定 6 次,并计算平均值。

1.6 统计分析方法

目前药检系统广泛采用 Z 比分数作为实验室比对和能力验证判定的参考指数。 $Z = (\text{结果} - \text{指定值}) / \text{能力评定标准差}$ 。 $|Z| \leq 2$ 时,结果为满意; $2 < |Z| < 3$ 时,结果判为可疑; $|Z| \geq 3$ 时,结果为不满意^[13]。

在进行 Z 值计算时,指定值和能力评定标准差有多个获得途径,如参加者、专家实验室等。在这里采用专家实验室获得参数的方法,其中能力评定标准差由精密度实验得出,组织专家实验室进行溶液 pH 测定,选取包括能力验证检测水平在内 5 个水平,对 8 个专家实验室进行精密度实验^[11],采用数理统计的方法,对实验数据处理,计算重复性标准差 S_r 和再现性标准差 S_R ,根据 ISO 13528(2015)第 8.5 节中规定^[15],计算, $\sigma_{pt} = \sqrt{S_R^2 - S_r^2(1 - 1/n)}$,确定能力验证标准差(σ_{pt})。

1.7 统计分析软件

在均匀性检验和稳定性检验中,使用 JMP 软件对数据进行处理。在离群值判别和 Z 值计算时使用 EXCEL 软件。对统计结果作图分析时,使用 SPSS 软件。

2 结果

2.1 均匀性检验结果

以水平 1 为例,采用 JMP 软件对数据进行分析,结果见图 2 和表 1。

此时, F 值对应的显著性水平为 0.639 2, 大于 0.05, 样品是均匀的。以相同方法, 对水平 2 的样品进行均匀性检验, 结果均符合规定。

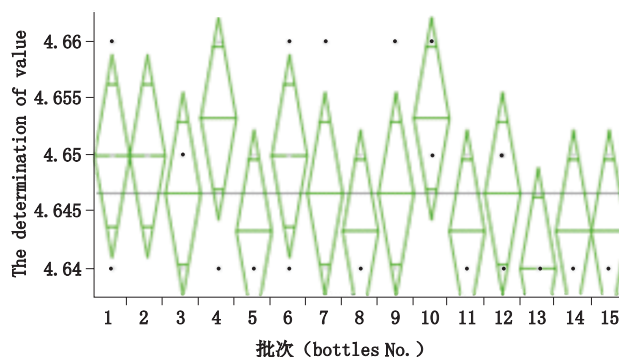


图 2 JMP 均匀性检验结果

Fig. 2 Results of homogeneity of samples

表 1 JMP 均匀性检验结果

Tab. 1 Results of homogeneity of samples

	自由度 (degree of freedom)	平方和 (sum of square)	均方 (mean square)	F 比 (F ratio)	概率 > F (probability > F)
瓶号 (bottle No.)	14	0.000 666 67	0.000 048	0.824 2	0.639 2
误差 (error)	30	0.001 733 33	0.000 058		
校正总和 (correction combined)	44	0.002 400 00			

2.2 稳定性检验结果

以水平 1 样品, 在 80% RH 条件下的稳定性检验为例, 使用 JMP 软件进行, 结果见图 3 和表 2。

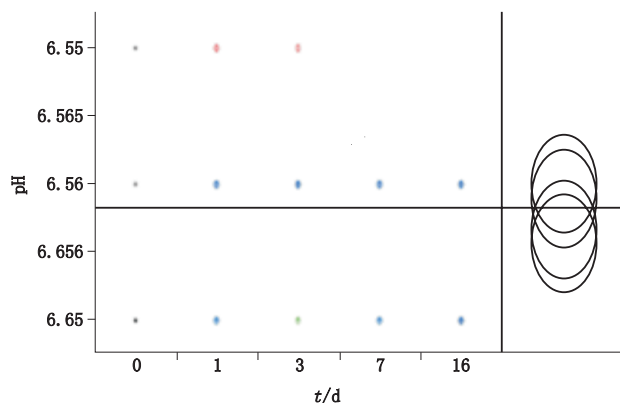


图 3 稳定性检验结果

Fig. 3 Results of stability of samples

LSD 阈值矩阵中均为负值, 表明各对均值无显著性差异。可得出结论, 水平 1 的样品的 80% RH 条件下, 稳定性能满足要求。使用 JMP 对水平 1 样品在 60 ℃, 和水平 2 样品在 80%RH 与 60 ℃, 这 3 种情况进行分析, 结果均符合要求。

2.3 精密度

2.3.1 专家实验室原始数据 汇总 8 个实验室的测量结果, 未发现明显不规则的测量数据。

2.3.2 一致性检验 对测试结果的一致性和离群值进行检查。计算各水平和实验室的曼德尔 h 统计量和 k 统计量, 结果见图 4。由图可知, 存在数个超过 h 临界值 2.06 和 k 临界值 1.97 的数据, 可能存在歧离值或离群值, 用数值法进行进一步检验。

表 2 稳定性检验阈值矩阵

Tab. 2 Threshold matrix of stability of samples

Abs (Dif) - LSD	0	3	1	14	7
0	-0.005 02	-0.005 02	-0.003 91	-0.001 69	-0.000 58
3	-0.005 02	-0.005 02	-0.003 91	-0.001 69	-0.000 58
1	-0.003 91	-0.003 91	-0.005 02	-0.002 80	-0.001 69
14	-0.001 69	-0.001 69	-0.002 80	-0.005 02	-0.003 91
7	-0.000 58	-0.000 58	-0.001 69	-0.003 91	-0.005 02

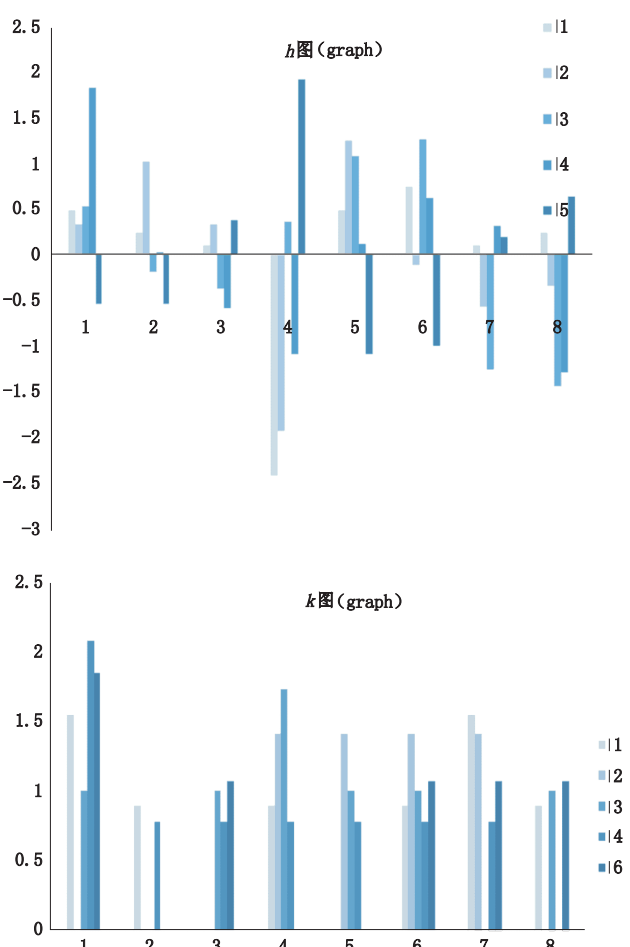


图4 按实验室进行分组的实验室间一致性曼德尔 *h* 图和 *k* 图
Fig. 4 Mandel's *h* and *k* graph for the consistency grouped by laboratories

2.3.3 柯克伦检验 柯克伦检验检验离群值,是对实验室内变异的检验。计算得到的柯克伦统计量 *C* 列于表 3。

表 3 柯克伦检验结果

Tab. 3 Results of Cochran test		
水平 <i>j</i> (level <i>j</i>)	<i>C</i>	检验类型 (test type)
1	0.311	柯克伦检验统计量 Cochran test statistics
2	0.250	
3	0.368	
4	0.538	
5	0.429	
歧离值 (straggling value)	0.561	柯克伦检验临界值
离群值 (outliers)	0.664	Cochran test threshold

没有歧离值,说明各参加实验室的室内方差均满足规定。柯克伦准则是单侧离群值检验,所验证的为 1 组标准差中的最大值。小的标准差值可能很大程度上受原始数据修约程度的影响,并不可靠。另外,似乎并没有理由拒绝一个比其他实验室精密度都要高的实验室的数据^[12]。因此柯克伦检验是合理的。

2.3.4 格拉布斯检验 格拉布斯检验主要是对实验室内变异的检验。将格拉布斯检验应用于单元平均值。计算得到的格拉布斯统计量列于表 4。

表 4 格拉布斯检验结果

Tab. 4 Results of Grubbs test

水平 <i>j</i> (level <i>j</i>)	单个低值 (single low value)	单个高值 (single high value)	2 个低值 (two low values)	2 个高值 (two high values)	检验类型 (test type)
1	2.414	0.758	—	—	格拉布斯检验统计量 (Grubbs test statistics)
2	1.928	1.225	0.275	0.508	
3	1.392	1.225	0.330	0.438	
4	1.279	1.821	0.470	0.321	
5	1.080	1.920	0.590	0.256	
歧离值 (straggling value)	2.126	2.126	0.110 1	0.110 1	格拉布斯检验临界值 (Grubbs test threshold)
离群值 (outliers)	2.274	2.274	0.056 3	0.056 3	

其中水平 1 的单个低值大于临界值,为离群值,将该水平最低值剔除,再次进行格拉布斯检验,无离群值。剔除水平 1 中的最低值单元,即实验室 4 在水平 1 的测量结果,其他数据予以保留。

2.3.5 参数计算 根据格拉布斯检验,剔除异常值后的实验数据,按 GB/T 6379.2—2004 进行统计分析。对每个水平 $\hat{m}$,计算重复性方差 S_r 和再现性方差 S_R ,结果见表 5。

表 5 不同水平 pH 测定的 $\hat{m}$, S_r 和 S_R 值Tab. 5 The results of $\hat{m}$, S_r and S_R of each level for pH

determination				
水平 j (level j)	p	$\hat{m}$	S_r	S_R
1	7	3.78	0.007	0.008
2	8	4.66	0.004	0.015
3	8	6.88	0.006	0.020
4	8	9.22	0.007	0.034
5	8	9.99	0.005	0.037

2.3.6 测量精密度与检测水平 m 的相关性 根据表 4, m 与重复性标准差 S_r 没有明显的依赖关系。随着 m 值的加大显然再现性标准差 S_R 也在增加, 所以可能可以建立再现性标准差与 m 之间的某种关系, 借助 Excel 对 S_R 与 m 的关系进行线性和对数 2 种拟合, 得到的趋势线与回归方程见图 5, 其中线性回归的相关系数 $r=0.988$, 对数回归的相关系数 $r=0.967$ 。

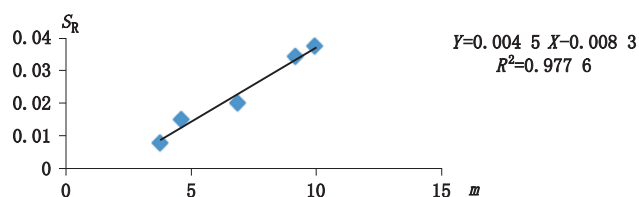


图 5 线性拟合函数关系图

Fig. 5 Linearity fitting graph

对于再现性而言, 2 条曲线拟合的都较好。但是拟合出的曲线与 y 轴的截距均为负值, 不符合 GB/T 6379—2004《测量方法与结果的准确度(正确度与精密度)》的第 2 部分《确定标准测量方法重复性与再现性的基本方法》中第 7.5.1 节中对于函数关系的要求, 不予采用。同时考虑到标准差的绝对值较小, 采用这些量的平均值作为重复性标准差和再现性标准差的终值。

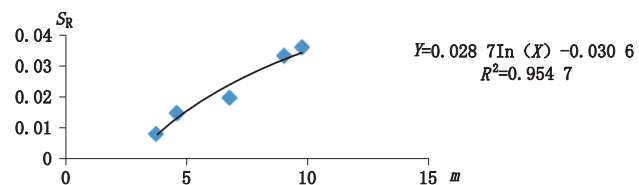


图 6 对数拟合函数关系图

Fig. 6 Logarithmic fitting graph

2.4 能力验证结果

2.4.1 能力评定标准差确定 鉴于测量精密度与检测水平 m 之间不存在相关性, 取各水平的平均值作

为测量方法精密度, 由此溶液 pH 测定的重复性标准差 $S_r=0.006$, 再现性标准差 $S_R=0.023$ 。

根据 ISO 13528(2015) 中第 8.5 节规定, 计算 $\sigma_{pt} = \sqrt{S_R^2 - S_r^2 (1 - 1/n)} = 0.03$

2.4.2 指定值确定 采用专家实验室测试结果的平均值作为指定值。对于水平 1, 指定值为 4.66; 水平 2, 指定值为 9.22。

2.4.3 评价结果 根据 $Z = (\text{结果} - \text{指定值}) / \text{能力评定标准差}$, 对提交结果的 251 家实验室进行数据统计。将此次能力验证的结果用象限散点图表示, 可以直观地展示数据的分布情况(见图 7)。

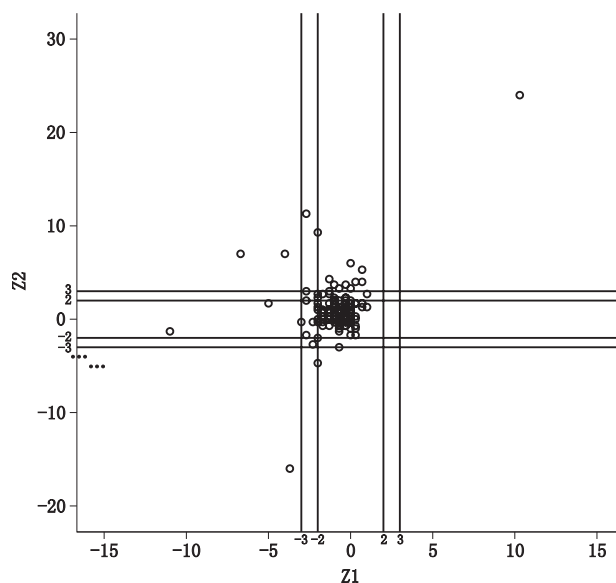


图 7 实验室 Z 比分数散点图

Fig. 7 The scatter diagram of laboratory Z score

检测水平 1 满意结果为 238 家, 占参加实验室的 94.82%; 可疑结果为 6 家实验室, 占参加实验室的 2.39%。检测水平 2 满意结果为 223 家, 占参加实验室的 88.84%, 可疑结果为 9 家, 占参加实验室的 3.58%。

各实验室检测项目中有 1 个水平不满意即判定此实验室评价结果为不满意, 剩余实验室中有 1 个水平为可疑即判定此实验室评价结果为可疑, 其余为满意结果。实验室总的满意率为 86.45%, 可疑率为 4.78%, 不满意率为 8.76%。

3 讨论

在能力验证评价时采用了选择专家实验室进行前期实验, 评价出 pH 测定检验能力的重复性和再现性标准差, 进而获得能力验证评价参数的方法。这种方式不同于目前国内外较常使用的以中位值和

NIQR 为评价参数的稳健性方法,在这种方法中评价参数独立于参加者结果,可以有效地防止整体样本数据偏离带来的评价误差,也在一定程度上减少了实验室串通对结果评价的影响。

本次能力验证中,尝试性的采用了一些新的统计软件。在均匀性检验和稳定性检验中,使用 JMP 软件对数据进行处理。JMP 算法源于 SAS(全球最大的统计学软件公司),是 SAS 推出的交互式可视化统计发现软件。在医药领域,FDA 对于药企申报的新药报告中的统计分析部分,只接受用 SAS 和 JMP 分析得出的统计结果。JMP 的统计分析是权威可靠的。该软件以简单的方式实现复杂高效的分析,无需学习统计原理,能解决实际问题,非常适合非统计专业背景的数据分析人员使用。

将 Z 比分数反映在散点图上,在简化了尤登图的复杂计算的同时,依然可以直观地展现数据的分布情况,比如实验室代码为 438 的实验室结果位于第一象限,其 2 份样本的 Z 值均远大于 3,实验室内存在明显的正偏差,说明该实验室内 pH 检测能力体系中存在系统性的正误差,应该从仪器状态、校正溶液、电极情况等硬件方法加以审核。

本轮能力验证结果中反映出有多个实验室是其中 1 个水平的测量结果不满意导致的不满意结果,表明其检测水平的稳定性欠佳,更应持续参加能力验证项目,提高准确率。

参考文献

- [1] ISO/IEC 17043:2010,能力验证提供者认可准则[S]. 2010:3. 7
ISO/IEC 17043:2010 General requirements for proficiency testing [S]. 2010:3. 7
- [2] BELK WP, SUNDERMAN FW. A survey of the accuracy of chemical analyses in clinical laboratories[J]. Am J Clin Pathol, 1947, 17(11): 853
- [3] SUNDERMAN FW. The history of proficiency testing/quality control [J]. Clin Chem, 1992, 38(7): 1 205
- [4] LAESSIG RH, EHRMEYER SS. Proficiency testing programs—promises, progress, and problems. A 40-year prospective[J]. Arch Pathol Lab Med, 1988, 112(4): 329
- [5] SUNDERMAN FW. Forty-five years of proficiency testing[J]. Ann Clin Lab Sci, 1991, 21(2): 143
- [6] 刘小娟, 黄亨建, 李萍. 以 CAP 能力验证提高临床生化检验质量[J]. 检验医学, 2005(5): 470
LIU XJ, HUANG HJ, LI P. Improving quality of clinical laboratory through taking part in College of American Pathologists proficiency testing[J]. Lab Med, 2005(5): 470
- [7] 王厚芳, 孙芾, 柳杰. 通过 CAP 能力验证提高血液寄生虫检查质量[J]. 实验与检验医学, 2006, 24(6): 495.
WANG HF, SUN F, LIU J. Improving quality of blood parasite check through taking part in College of American Pathologists proficiency testing[J]. Exper Lab Med, 2006, 24(6): 495
- [8] 柯培峰, 梁铮, 黄迪, 等. 化学发光检测系统参加 CAP 校准检验活动结果的分析[J]. 实用医学杂志, 2011, 27(12): 2254
HE PF, LIANG Z, HUANG D, et al. Analysis of the results of CAP calibration verification in the chemiluminescence detection system [J]. J Pract Med, 2011, 27(12): 2254
- [9] CNAS-RL02:2010,能力验证规则[S]. 2010:5
CNAS-RL02:2010, Rules for Proficiency Testing[S]. 2010:5
- [10] 宋桂兰. 能力验证工作历史回顾[J]. 中国计量, 2009(2): 35
SONG GL. The review of ability verification work history[J]. China Metrol, 2009(2): 35
- [11] 毛敏, 项新华, 肖新月. 2013 年食品药品系统实验室能力验证结果分析[J]. 中国药事, 2015(29): 1145
MAO X, XIANG XH, XIAO XY. Proficiency testing results of food and drug laboratories in 2013[J]. Chin Pharm Aff, 2015(29): 1145
- [12] CNAS-G03:2006,能力验证样品均匀性和稳定性评价指南[S]. 2015:4
CNAS-G03:2006, Guidance on Evaluating the Homogeneity and Stability of Samples Used for Proficiency Testing[S]. 2015:4
- [13] CNAS-GL02:能力验证结果的统计处理和评价指南[S]. 2015:4. 4
CNAS-GL02: Guidance on Statistic Treatment of Proficiency Testing Results and Performance Evaluation[S]. 2015:4. 4
- [14] 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[S]. 2005:7. 3
GB/T 6379.2—2004, Accuracy(trueness and precision) of measurement methods and results—Part2: Basic method for the determination of repeatability and reproducibility of a standard measurement method[S]. 2005:7. 3
- [15] ISO 13528:2015,实验室间比对能力验证用统计方法[S]. 2015:8. 5
ISO 13528:2015, Statistical methods for use in proficiency testing by interlaboratory comparison[S]. 2015:8.5

(本文于 2017 年 12 月 6 日修改回)