

文章编号: 1001-764X(2013) 12-881-04

• 临床检验方法性能评价 •

ISO 15189: 2012 与临床检验定量检测方法确认和性能验证

康凤凤, 王治国(北京医院, 卫生部临床检验中心, 北京 100730)

摘要: ISO 15189: 2012 对技术要求的检验过程分别单独列出了对检验程序确认和验证、测量不确定度的要求, 并将其改为强制要求。本文参考相关文件, 结合我国实际情况, 扼要介绍方法性能评价关键指标的基本概念及不同检测系统的实施方案。对配套系统, 实验室应进行验证实验, 至少包括精密度、正确度和可报告范围; 对非配套系统, 实验室应进行确认研究, 至少包括正确度、精密度、可报告范围及生物参考区间等。以期为临床实验室更合理地评价和应用检测方法提供参考。

关键词: 确认; 验证; 精密度; 正确度; 可报告范围; 参考区间

中图分类号: R446.1

文献标志码: A

ISO 15189: 2012^[1] 较 2007 年版对技术要求的检验程序提出了更详细的要求。ISO 15189: 2007^[2] 5.5.2 指出实验室只能使用已确认的适用于预期用途的检验程序, 确认程序应尽可能充分, 以满足给定用途或某特定领域的应用要求。ISO 15189: 2012 特别具体列出了 5.5.1 部分“检验程序的选择、验证和确认”, 包括检验程序的验证、检验程序的确认及测量量值的测量不确定度。但众所周知, ISO 15189 为总纲性的准则文件, 具体该如何实施方法确认和性能验证对很多实验室而言仍感疑惑。国内已有很多关于方法确认和性能验证的文献, 介绍方法多样性的同时也为临床工作者增加了负担。本文结合国内外最新文件, 阐述临床实验室方法确认和性能验证的方法及存在的问题。

1 基本概念

1.1 检测系统 检验系统是完成一个检测项目需要的仪器、试剂、校准品、操作程序、操作人员等的组合^[3]。更广义地说, 为保证检测质量, 还应将检测系统配套的标本采集管、配套离心机、电源等也纳入检测系统。《医疗机构临床实验室管理办法》(卫医发〔2006〕73 号)^[4] 第二十四条要求医疗机构临床实验室应当保证检测系统的完整性和有效性, 对需要校准的检验仪器、指标和对临床检验结果有影响的辅助设备定期进行校准。

1.2 随机误差(random error) 在重复测量中按不可预见方式变化的测量误差的分量^[5]。通常随机误差的参考量值是对同一被测量无穷多次重复测量得到的均值。

1.3 系统误差(systematic error) 在重复测量中保持不变或按可预见方式变化的测量误差的分量^[5]。系统误差的参考量值是真值, 或者是测量不确定度可忽略不计的测量标准测得的值, 还可是约定量值。

1.4 验证(verification) ISO 15189: 2012^[1] 定义: 通过提供客观证据对规定要求已得到满足的认定。其中 5.5.1.2 检验程序验证, 强调未经修改而使用的已确认的检验程序在常规使用前, 应经过实验室的独立验证。实验室应从制造商或方法开发者处获得相关信息, 以确定检验程序的性能特征。实验室进行的独立验证, 应通过获取客观的证据(以性能特征形式)证实检验程序的性能与其声明相符。验证过程证实的检验程序的性能指标, 应与检验结果的预期用途相关。GB/T 22576-2008^[6] 明确指出, 临床实验室使用厂家已严格评估的检验方法或试剂盒之前, 应验证相关分析性能以证实在本实验室能达到厂家声明的性能指标。卫生行业标准 WS/T 420-2013^[7] 指出临床实验室应对所用的检验方法或商品试剂盒进行验证。CNAS-CL38^[8] 要求性能验证内容至少应包括正确度、精密度和可报告范围。

1.5 确认(validation) ISO 15189: 2012^[1] 定义: 通过提供客观证据对特定的预期用途或应用要求已得到满足的认定。其中 5.5.1.3 检验程序的确认: 实验室应对以下来源的检验程序进行确认: (a) 非标准方法; (b) 实验室设计或制定的方法; (c) 超出预期范围使用的标准方法; (d) 修改过的确认方法。方法确认应尽可能全面, 并通过客观的证据(以性能特征形式)证实满足检验预期用途的特定要求。而我国卫生行业标准 WS/T 420-2013^[7] 指出以下两

作者简介: 康凤凤, 1988 年生, 女, 硕士研究生, 研究方向为临床实验室质量管理。

通信作者: 王治国, E-mail: zgwang@nccl.org.cn。

种情况下临床实验室需进行确认试验: (1) 临床实验室建立的新检验方法或试剂盒; (2) 对我国食品药品监督管理局 (FDA) 批准的检验方法或试剂盒进行重要修改。CNAS-CL38^[8] 要求性能确认研究至少应包括正确度、精密度、可报告范围及生物参考区间等。

2 性能特征

2.1 精密度 (precision) (1) 概念: 在规定条件下, 对同一或类似被测对象重复测量所得示值或测得值间的一致程度^[5]。规定条件可以是重复性测量条件、期间精密度测量条件或复现性测量条件。精密度通常无法衡量, 而以不精密度表达, 常用标准差、方差或变异系数表示, 是对随机误差的衡量。(2) 验证方案: 对配套系统, 精密度验证可参照 CLSI EP15^[9] 和 WS/T 420-2013^[7] 文件, 注意这必须建立在厂家已按 CLSI EP5 文件对检测系统的精密度进行确认的基础上。具体验证方案: 连续 5 d, 每天检测 1 个分析批, 包括 2 个浓度水平, 分别对每个浓度水平的样品重复检测 3 次。将评价得到的不精密度与厂家声明比较, 证实是否符合厂家声明。Westgard 法^[10] 验证精密度与 EP 文件不同, 该方法较灵活, 临床实验室可根据自身的实验室检测水平、人力、财力等, 选择合适的检测数, 但最低的研究方案应包括: 根据医学决定水平选择至少 2 个浓度水平样品, 在 1 d 或多天内获得至少 20 个检测数。如果 20 个检测值在 1 批或 1 d 内获得, 可估计短期不精密度, 而在 20 d 以上获得可估计长期不精密度。用方差分析判断是否符合厂家说明。国内外实验室对厂家配套系统的验证研究可供参考^[11-12]。(3) 确认方案: 对非配套系统, 或者厂家未提供精密度声明, 或是实验室自己建立的方法, 则可按照 CLSI EP5^[13] 文件确认方法的精密度。具体方案为: 至少两个浓度水平的样品, 每天重复检测 2 次, 连续 20 d。精密度评价标准可参考卫生行业标准 WS/T 403-2012^[14]: 临床生物化学检验常规项目分析质量指标, 该行业标准是在生物学变异的基础上建立的质量规范, 在质量规范层次模型中处于第二层次。将评价的不精密度与行业标准中的允许不精密度比较, 判断方法的可接受性。对于部分检测项目, 可能没有行业标准, 可用方法决定图判断^[10]。Højskov 等^[15] 对自建检测系统的确认研究可供参考。而 Westgard^[10] 推荐可接受性准则为: 短期 (批内或日内) 不精密度小于 0.25 倍允许总误差 (allowable total error, TEa), 长期 (日间或实验室内) 不精密度小

于 0.33TEa, TEa 可来自我国室间质量评价要求。

2.2 正确度 (trueness) (1) 概念: 无穷多次重复检测所得量值的平均值与一个参考量值间的一致程度。通常用“偏移”表达, 是对系统误差的衡量。选择一个合适的参考对正确度评价很重要, 最好采用参考方法^[16-17]。但由于成本问题, 实验室通常采用已经使用的常规方法进行比对试验^[18]。注意此时评价的并非方法真正的偏移, 而是评价方法和比对方法的偏差或差值。(2) 验证方案: 对配套系统, 实验室无需再按照 CLSI EP9 文件执行正确度确认, 而按照 CLSI EP15^[9] 和 WS/T 420-2013^[7] 文件执行正确度验证方案, 包括两种方式: 使用患者标本和使用参考材料。采用患者标本推荐分别用评价方法和比对方法单次分析 20 份患者标本, 计算偏移置信区间并与厂家声明比较。另一种方法是采用有证参考材料^[19], 后者可来源于室间质量评价正确度验证计划、厂家提供的正确度验证计划的质控品、其他第三方供应商提供等。至少选择两个浓度水平的参考材料, 对每个参考材料分别在 3~5 个不同批次中重复测量 2 次, 计算均值与指定值比较, 估计偏移, 验证厂家声明。目前卫生部临床检验中心已开展代谢物/总蛋白质、脂类、糖化血红蛋白正确度验证等计划, 参加实验室近 200 家。通过参加此类正确度验证计划, 临床实验室可直接从反馈报告中获取相应指标的偏移估计。我中心也提供了部分检测指标的标准物质, 可登陆我中心网站 (www.ncccl.org.cn) 查看。(3) 确认方案: 对非配套系统, 实验室需要按照 CLSI EP9^[20] 文件执行正确度确认试验, 分别用评价方法和比对方法分析至少 100 个患者标本。比对方法可以是实验室当前采用的测量程序、厂家声明中使用的测量程序或认可的参考测量程序。如果不采用参考测量程序, 其他比对方法可能会有干扰物质的影响和基质效应。值得注意的是, 分析比对结果时常采用回归分析法, 最常见的是一般线性回归模型。但该回归模型的重要前提是假定比对方法不精密密度为 0, 而事实上比对方法不精密度通常不能忽略。此时可采用加权一般线性回归或 Deming 回归分析, 后者考虑了两种方法的不精密度。正确度评价质量要求可参考 2012 年颁布的卫生行业标准 WS/T 403-2012^[14]。对于部分没有行业标准的检测指标, 可用方法决定图判断^[10]。

2.3 可报告范围 (reportable range) (1) 概念: 按照 CLSI 一致化术语数据库^[21], 可报告范围指测量仪器的误差在预期规定范围内的被测量值的集合。

可报告范围等同于测量范围 (measurement range), 后者定义为实验室可建立或验证仪器、试剂或检测系统测量响应的准确性的检测结果的范围。另外还有一个术语称为临床可报告范围 (clinically reportable range), 定义为方法可采用标本的稀释、浓缩或其他预期处理用于延长直接的分析测量范围下的分析物值的范围, 其范围一般较测量范围宽, 是考虑方法检测限和预处理流程后的测量范围的延伸。(2) 验证方案: CLSI EP6^[22] 文件提供了估计可报告范围的指南, 推荐分析 5 个以上的不同浓度水平的样品, 可采用高浓度和低浓度样品按不同比例混合得到系列浓度样品, 后者应覆盖整个预期测量范围并包括临床有意义的浓度水平。每个水平的样品重复检测 2 次。(3) 确认方案: CLSI EP6 文件表明新方法的报告范围的建立需采用 7~11 个浓度的样品, 最好是较预期测量范围宽 20%~30% 的浓度范围。根据预期不精密度对样品重复检测 2~4 次。

临床实验室也可参考卫生行业标准 WS/T408-2012《临床化学设备线性评价指南》^[23], 推荐验证实验采用 4~6 个浓度水平, 确认实验采用 9~11 个浓度水平, 每个样品重复检测 3~4 次。

2.4 参考区间 (reference interval) (1) 概念: 又称为参考范围、正常范围等, 是参考值分布的中间 95% 区间^[1]。部分检测指标只有一个参考限具有临床意义, 如参考上限, 则参考区间为 0~参考上限。(2) 验证方案: CLSI C28^[24] 文件建议采用 20 个患者样品来验证厂家提供的参考区间。我国卫生行业标准 WS/T 402-2012^[25] 推荐 3 种方法验证参考区间: (a) 实验室判断实际情况与厂家提供的分析程序、参考区间估计方法、人群分布、人口统计学资料等一致时, 可直接使用; (b) 从本地参考人群中筛选 20 个参考个体验证; (c) 增加样品量至 60 验证参考区间, 可增加统计学原理发现人群差异的能力。(3) 确认方案: 如果是实验室建立的新方法或对配套系统做出重要改变时, 需要重新建立新的参考区间, 此时需至少纳入 120 个个体。注意有些检测物质, 如肌酐、红细胞计数等的参考区间受年龄和性别的影响, 应考虑分组建立参考区间, 并对每一个分组纳入 120 个个体。验证或建立参考区间的关键在于理解所有影响检测结果的分析前变量, 并根据关键变量选择健康人群合理分组。

2.5 检测限 (limit of detection) 最早用术语“分析灵敏度”来描述检测下限, 同时另一术语为“功能灵敏度”, 即测量程序长期不精密度 (CV) 为 20% 时

应的浓度水平, 该定义基于精密度曲线。由于“分析灵敏度”与术语“检出限”或“检测低限”在一定程度上可互换, 而“功能灵敏度”又易与术语“临床灵敏度”、“阈值”等混淆, CLSI EP17-A2^[26] 文件推荐用新术语描述检测下性能: 空白限 (limit of blank, 空白样品的系列结果中的最大值)、检出限 (limit of detection, 方法可检出的最低被测物浓度) 和定量限 (limit of quantitation, 能可靠检出分析物且检测结果的不确定度满足实验室既定目标的实际浓度)。该文件推荐至少检测 60 个空白样品和低值样品来建立空白限和检测限, 而验证厂家声明时可仅采用 20 个检测。温冬梅等^[27] 对甲胎蛋白的研究可供参考。

2.6 干扰 (interference) 干扰是指被测物质浓度因样品特性或其他成分的影响而出现的临床显著性偏差, 评价的是由样品引起的特异性偏差。干扰实验可参考 CLSI EP7^[28] 文件及 WS/T 416-2013《干扰实验指南》。通常选择 2 个医学决定水平的被测量浓度, 或者根据参考范围的高低限, 分别加入不同的高浓度、不同类型的可能干扰物进行筛查实验, 与正常样品结果比较, 通过 *t* 检验明确差异有统计学意义的干扰物质。将干扰物高浓度实验样品与低浓度实验样品按不同比例混合, 建立介于两者浓度间的一系列实验样品。样品重复检测 3 次以排除方法精密度的影响。通过回归分析明确干扰物剂量效应, 以确定有临床意义出现的干扰物浓度水平。国内也有学者对临床化学抗干扰性能进行了评估^[29]。

3 临床性能特征

临床检验结果最终目的是用于患者的临床诊疗, 因此检验方法的评价还应包括临床性能的评价^[30], 后者主要包括诊断或鉴别诊断的准确度、决策性或有效性。检验方法临床评价的本质是检验方法结果与患病与否之间的关系, 这种能力可用敏感性和特异性衡量。敏感性指金标准诊断有病的全部病例中, 被评价方法结果为阳性的病例占全部有病病例的比例, 即敏感性 = [真阳性 / (真阳性 + 假阴性)] × 100%; 特异性指金标准诊断全部无病病例中, 被评价检验方法结果为阴性的病例所占全部无病病例的比例, 即特异性 = [真阴性 / (假阳性 + 真阴性)] × 100%。临床性能评价过程首先应确定具体的临床问题, 以此为基础选择合适的样品。

4 测量不确定度

目前关于测量不确定度是否应纳入临床实验室

检测结果的争论愈发激烈^[31]。ISO 15189: 2007^[2]指出,实验室应明确性能参数,包括线性、精密度、以测量不确定度表示的准确性等。ISO 15189: 2012^[1]中没有提到用测量不确定度表示准确性,而是单独要求实验室明确测量程序的测量不确定度。测量不确定度与误差不同,前者是表征赋予被测量值的分散性,后者是表示测量结果偏离真值的程度。中国合格评定国家认可委员会颁布的《医学实验室—测量不确定度的评定与表达》^[32]推荐了多种确定实验室检测结果的测量不确定度的方法。鉴于临床实验室实际工作,可选择利用室内质控和室间质评数据计算检测结果的测量不确定度。

5 参考文献

- [1] ISO 15189. Medical laboratories—Requirements for quality and competence [S]. International Organization for Standardization, 2012.
- [2] ISO 15189. Medical laboratories—Particular requirements for quality and competence [S]. International Organization for Standardization, 2007.
- [3] 冯仁丰. 临床检验质量管理技术基础 [M]. 第 2 版. 上海: 上海科学技术文献出版社, 2007: 37-38.
- [4] 卫医发(2006)73 号. 医疗结构临床实验室管理办法 [S]. 中华人民共和国卫生部, 2006.
- [5] JJF 1001-2011. 通用计量术语及定义 [S]. 国家质量监督检验检疫总局, 2011.
- [6] GB/T 22576-2008. 医学实验室质量和能力的专用要求 [S]. 中国国家标准化管理委员会, 2010.
- [7] WS/T 420-2013. 临床实验室对商品定量试剂盒分析性能的验证 [S]. 中华人民共和国国家卫生和计划生育委员会, 2013.
- [8] CNAS-CL38. 医学实验室质量和能力认可准则在临床化学检验领域的应用说明 [S]. 中国合格评定国家认可委员会, 2012.
- [9] CLSI EP15-A2. User verification of performance for precision and trueness; approved guideline—second edition [S]. CLSI, 2005.
- [10] 王治国. 临床检验方法确认与性能验证 [M]. 北京: 人民卫生出版社, 2008.
- [11] Knudsen CS, Nexø E, Højkskov CS, *et al.* Analytical validation of the Roche 25-OH vitamin D total assay [J]. Clin Chem Lab Med, 2012, 50(11): 1965-1968.
- [12] 徐建辉, 王惠民, 苏建友, 等. 商品定量试剂盒分析性能的验证试验 [J]. 临床检验杂志, 2010, 28(6): 467-468.
- [13] CLSI EP5-A2. Evaluation of precision performance of quantitative methods; approved guideline—second edition [S]. CLSI, 2004.
- [14] WS/T 403-2012. 临床生物化学检验常规项目分析质量指标 [S]. 中华人民共和国卫生部, 2012.
- [15] Højkskov CS, Heickendorff L, Møller HJ. High-throughput liquid-liquid extraction and LCMSMS assay for determination of circulating 25(OH) vitamin D3 and D2 in the routine clinical laboratory [J]. Clin Chim Acta, 2010, 411(1-2): 114-116.
- [16] 陈宝荣, 孙慧颖, 邵燕, 等. 用 IFCC 参考方法评价 11 个常规系统测量血清 α -淀粉酶结果的正确性 [J]. 临床检验杂志, 2011, 29(4): 309-313.
- [17] 孙慧颖, 邵燕, 胡滨, 等. 7 种血清葡萄糖常规检测系统测量结果的正确度评价 [J]. 临床检验杂志, 2013, 31(7): 542-544.
- [18] Bossuyt PM, Irwig L, Craig J, *et al.* Comparative accuracy: assessing new tests against existing diagnostic pathways [J]. BMJ, 2006, 332(7549): 1089-1092.
- [19] 孙慧颖, 陈宝荣, 邵燕, 等. 人血清标准物质在改进临床实验室 T-Bil 测量结果正确度中的价值研究 [J]. 临床检验杂志, 2013, 31(4): 299-302.
- [20] CLSI EP9-A3. Measurement procedure comparison and bias estimation using patient samples; approved guideline—third edition [S]. CLSI, 2013.
- [21] CLSI. Harmonized terminology database [DB]. <http://login.clsi.org/HTDatabase.cfm>. accessed on July 2008.
- [22] CLSI EP6-A. Evaluation of the linearity of quantitative measurement procedures: a statistical approach; approved guideline [S]. CLSI, 2003.
- [23] WS/T 408-2012. 临床化学设备线性评价指南 [S]. 中华人民共和国卫生部, 2012.
- [24] CLSI C28-A3c. Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline—third edition [S]. CLSI, 2008.
- [25] WS/T 402-2012. 临床实验室检验项目参考区间的制定 [S]. 中华人民共和国卫生部, 2012.
- [26] CLSI EP17-A2. Evaluation of detection capability for clinical laboratory measurement procedures; approved guideline—second edition [S]. CLSI, 2012.
- [27] 温冬梅, 张秀明, 王伟佳, 等. 化学发光免疫法检测 AFP 的空白限、检出限和定量检测限的建立与评价 [J]. 临床检验杂志, 2010, 28(6): 469-471.
- [28] CLSI EP7-A2. Interference testing in clinical chemistry; approved guideline—second edition [S]. CLSI, 2005.
- [29] 文江平, 郭拥军, 鲁辛辛, 等. 临床实验室认可中临床化学干扰性能的系统评估 [J]. 临床检验杂志, 2008, 26(3): 216-218.
- [30] Bossuyt PM, Reitsma JB, Linnet K, *et al.* Beyond diagnostic accuracy: the clinical utility of diagnostic tests [J]. Clin Chem, 2012, 58(12): 1636-1643.
- [31] 冯仁丰. 临床常规检验结果是否必须引入不确定度? [J]. 临床检验杂志, 2012, 30(10): 739-741.
- [32] TRL-001: 2012. 医学实验室—测量不确定度的评定与表达 [S]. CNAS, 2012.

(收稿日期: 2013-10-30)

(本文编辑: 王海燕, 陈维忠)