

临床实验室室间质量评价不同计划的结果评价探讨与研究

钟堃^a, 方遥^b, 王艳惠^c (北京医院/国家老年医学中心/中国医学科学院老年医学研究院)

a. 国家卫生健康委临床检验中心; b. 纪检监察室; c. 人事处, 北京 100730)

摘要: 在室间质量评价计划(能力验证)中, 对各参评实验室回报结果的处理和评价是最为重要的环节之一, 直接影响到室间质量评价计划的有效性, 这也是室间质量评价组织者和参评单位最常出现争议的环节。室间质量评价计划(external quality assessment, EQA)的结果和反馈报告应及时有效, 应包括(但不限于): 定量、半定量和定性项目所有参评实验室结果的统计分析、总体和不同分组中离群值的判断与剔除、单样本和多样本以及单参数和多参数的结果和分布的评价、指定值的确定、标准差的确定、图形显示, 以及在不同室间质量评价计划中多个样本和参数的整合信息等。室间质量评价计划结果的处理与评价方式一直在向着更加客观与合理的方向发展。

关键词: 室间质量评价; 统计分析方法; 结果评价与反馈

中图分类号: R446 文献标识码: A 文章编号: 1671-7414(2022)02-162-06

doi:10.3969/j.issn.1671-7414.2022.02.033

Discussion and Research on the Result Evaluation of External Quality Assessment Scheme for Clinical Laboratory

ZHONG Kun^a, FANG Yao^b, WANG Yan-hui^c (a. National Center for Clinical Laboratories;

b. Department of Supervision; c. Department of Human Resource, Beijing Hospital/National Center of Gerontology/Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, P.R. China, Beijing 100730, China)

Abstract: The work of handling data and evaluating results in external quality assessment (EQA) scheme is one of the most important parts, which will directly affect the performance of EQA scheme. The results of EQA scheme and participants' feedback reports should be timely and effective. The contents of the results and reports should include: the performances of quantitative, semi-quantitative and qualitative analysis, the elimination of outliers, the distribution evaluations of results of one or multiple parameters and/or one or multiple samples, determining the assigned values; determining the standard deviation, graphical presentation for one or more parameters or samples, combining information from different parameters and/or samples, etc. Staff should always pursue more objective and accurate methods and procedures for handling EQA results.

Keywords: external quality assessment; statistical analysis method; performance evaluation and result feedback

临床实验室的室间质量评价, 又称为外部质量评价(external quality assessment, EQA), 简称为室间质评, 不同国家或行业也称为能力验证(proficiency testing, PT)。常规的EQA是利用实验室间检验结果的比对来评估某实验室检验结果准确度的活动, 其目的是确保实验室检测能力的维持。EQA在国内和国际上都是一科学有效的评估手段^[1]。EQA可覆盖检验的全过程^[2]。其核心环节是对参评单位结果的评估及结果回报, 这是EQA组织者为参加用户提供的最为主要的反馈方式^[3]。如何对回报结果进行数据分析和统计学处理则是整个结果评价过程中最为重要的环节, 直接影响到EQA的有效性。国内EQA组织者多是通过每个项目单独评分, 之后再对检测能力进行评价, 从而在整体上给出实验室质量水平的改进情况^[4-6]。本文将对EQA定量、定性、半定量计划的统计分析、图形展示、多样本信息整合以及结果回报等进行探讨, 以期为

EQA提供者、参评用户, 广大实验室工作人员和质控人员提供信息和帮助。

1 室间质评定量结果的统计分析和处理

1.1 结果参数和分布评价 EQA提供者应对每轮次的回报结果进行分析和统计学处理, 得出如下参数: 结果数量/各分组结果数量、算数平均数、中位数、最大值、最小值、极差、标准偏差和变异系数等。美国病理家学会(CAP)建议每组最小样本数量应大于10, 国家卫生健康委员会临床检验中心建议每组数量应大于12。理论上结果应近似正态分布。而实际中可能存在多峰(multimodal)、正/负偏度(positive/negative skewness), 以及离群值(outlier)等。

多峰指在结果中有不同均值的分布存在, 而且在尚未找出原因的情况下, 对每个峰的结果群可考虑单独分组评估(peer grading)。正/负偏度是指数据不在均值周围对称, 均值在峰值的右边为正偏,

均值在峰值的左边为负偏。离群值一般远离大部分数据,可以认为获得该值的过程与其他数据不同。离群值可能是由于非分析性错误(extra-analytical error),比如记录或样本号错误等造成的,也称粗大误差(blunder)。EQA提供者可考虑通过数据转换来对正/负偏度进行修正,或其它统计学方法处理数据,比如使用核密度估计等方法来确定多峰,对于产生错误结果的原因可以充分的利用^[7-8]。

1.2 离群值(outlier)的鉴别及剔除 目前还没有一种可靠的方法能100%检出离群值,在检测离群值时常出现遮掩(masking)和无效(swamping)等情况。遮掩是指由于某个/某些离群值的存在而无法检测到其他的离群值,无效是指离群值检测的参数失效^[9-10]。常用的离群值检测方法有以下3种:Hampel(H法),Grubbs(G法)和Dixon(D法)。H法和G法是通过比较极端值与整体数据均值差异来进行判断。H法的基本步骤和原理是:计算 Z_i 值, $Z_i=[0.674\ 5(X_i-X_m)]/\text{MAD}$ (median absolute deviation,中位数绝对偏差), X_m 是组中各回报结果 X_i 的中位数,MAD是该批检测数据 X_i 减去中位数 X_m 差值之绝对值的中位数, Z_i 的绝对值 ≥ 3 则认为 X_i 是离群值。G法基本原理如下:一组中的回报结果按照从小到大排列,计算均值和标准差。然后计算均值与最大值和最小值的差值,确定一个差值最大的可疑值,计算可疑值的残差(可疑值与估计值的差值)与标准差的比值,查Grubbs临界值表来确定是否为临界值。D法则采用极差比的方法,设组内检测结果为顺序统计量 $X_1 < X_2 < \dots < X_n$ 。 X_i 服从正态分布时,D法在不同样本数量时统计量D的计算公式不同,有高端离群值和低端离群值两种检验,在确定显著性水平 α 后,根据D法的临界值表判断离群值。目前认为H法和G法相比D法更适用,而且G法能够检测少到6个数据样本的离群值^[11]。

EQA提供者还可以采取其他方式处理离群值。迭代法^[12]和标准化四分位间距法^[13]等稳健统计方法(robust statistics)可以保留离群值,因为它们是采用不同数据的权重不同来进行运算的^[14]。标准化四分位间距法是在保留所有数据的情况下计算总体均值和总体标准差的估计值:总体均值等于中位值;标准差等于标准化四分位间距(normalization inter quartile range, NIQR), $\text{NIQR}=0.741\ 3 \times \text{IQR}$ (inter quartile range,四分位距),等于高四分位数和低四分位数的差值。除此之外,还可以采用算数平均数或中位数除外3倍或者2倍标准差的方法来剔除离群值^[15]。尽管剔除的离群值结果不用于指定值(target)和样本标准差(s)等的计算,最终零

(0)评分应该体现在该回报结果所属用户的评估报告中。

1.3 指定值/靶值(target value)的确定 EQA中指定值/靶值(assigned value/target value)的确定可以参考ISO13528,一般可包括5种不同情况:定量配制值(formulation)–在特定样品配制(如制备、稀释)时确定;有证参考值(certified value)–根据定义的检测或测量方法确定(针对定量检测);参考值(reference value)–根据对EQA样品和可溯源到国家标准或国际标准的标准物质/标准样品或参考标准的并行分析、测量或比对来确定;专家实验室公议值(consensus values from expert laboratories)–专家级实验室的参与者(某些情况下可能是参考实验室)本身具有可证实的测定被测量的能力,并使用已确认的、有较高准确度的方法来进行检测,且该方法与常用方法有可比性;参加实验室的公议值(consensus values from participants)–利用参加单位检测的结果来进行确定^[12]。目前世界上大多数常规的EQA计划都是使用公议值(多为参加实验室的公议值)来作为指定值^[16]。而精准EQA计划的靶值来源于由参考实验室用参考方法测量的靶值,比如正确度验证计划等。

在任何的结果分布中,离群值都会对指定值的估计产生或多或少的的影响,这取决于离群值的大小和所采用的统计学方法。如果计算算数平均数作为指定值,则应在确认和剔除离群值后再进行计算。如果需要保留所有数据,则计算稳健平均值更为适用。对于稳健统计量,可以使用截点(breakdown point)和效能(efficiency)两个指标来进行评价。截点是指在其范围之外的数据不会影响稳健统计量。因此截点越大,越可能保留更多的离群值。效能则反映了估计值的不确定性。高截点和高效能是相悖的,算数平均值是高效能但截点很低。中位数广泛用于指定值估计,截点较高,效能较低。还可通过某些公式来计算指定值的估计值,设定参数 θ 是公式 $\sum_{i=1}^n |x_i - \theta|^p$ 数值最小的结果估计值,此处 X_i 是指 n 个数据点,也就是各参加者的结果, p 应在公式进行计算前定义^[17]。对于一个 p 值,只对应一个 θ 数值使公式运算结果最小。这个 θ 的数值称为最低效能估计。算数平均数是把 p 设定为2,而中位数是把 p 设定为1。算数平均数会由于离群值的影响而偏移很大,却有较高效能;而中位数的效能较低。这时可以寻找二者之间的折中值。可以把 p 设定为1.5,称为L1.5估计值。这个指定值的估计值既比中位数效能高又比算数平均数较少受到离群值的影响。

1.4 标准偏差的确定 EQA提供者有多种方法来

确定标准偏差,一般可分为两类:第一类包含了事先修正过的参数,可以是对参加实验室表现的预期、法律法规文件,或是一个小范围的精密度模型实验,比如 Horwitz 曲线分析^[18]等。如果有以往数据,标准差可以通过指定值来获得,比如基于指定值标准差数学关系估计的应用特征函数^[19]:

$$SD = \sqrt{\alpha^2 + \beta^2 \times (\text{指定值})^2}$$

这里 α 和 β 是用以往数据的非线性回归均值估计的。参数 α 从理论上解释了该项目低浓度时的标准偏差,参数 β 解释了高浓度的标准偏差并且当 β 较小或者浓度较高时接近变异系数。

第二类是基于回报结果的标准偏差估计。因为回报数据可能包含离群值,所以在计算标准偏差前应剔除离群值。EQA 提供者也可以使用稳健标准偏差。ISO 13528 标准中建议使用 Huber 的 M 估计量 H1.5 (算法 A) 来进行标准偏差的估计^[12]。还可以使用稳健 Q_n 估计量,数据中相同数据较多时会降低 Q_n 的可信度。还可以使用四分位间距计算,标准偏差的估计值等于四分位间距除以 1.349。

1.5 Q 评分和 Z 评分 定量项目一般可用两个评分准则:Q 评分和 Z 评分。

Q 评分指的是相对偏差,计算公式为:

$$Q \text{ 评分} = \frac{\text{实验室回报结果} - \text{指定值}}{\text{指定值}}$$

Q 评分与允许偏差比较来判断结果是否可接受。评价范围可以参考外部要求、分析质量规范、生物学变异等所制定的相应要求^[20-22]。

Z 评分是结果与指定值相差的标准偏差的倍数,计算公式为:

$$Z \text{ 评分} = \frac{\text{实验室回报结果} - \text{指定值}}{\text{标准偏差}}$$

若回报结果呈正态分布,Z 评分也会呈正态分布。对于正态分布来说 Z 评分绝对值在 2 和 3 以外的结果比例分别是 4.6% 和 0.27%。一般在 Z 评分绝对值小于 2 时认为结果可接受,在 2 与 3 之间怀疑该结果存在问题,而大于 3 则认为是不可接受的结果。

2 定性和半定量结果的统计分析和处理

对于定性和半定量结果,无法使用 Q 评分和 Z 评分。结果的评价取决于实验室是否给出了公议值或预期评价的结果。对于定性评价中的阳性样本,可以使用灵敏度 (sensitivity) 和阳性预测值 (positive predictive value) 进行评价。特异度 (specificity) 是阴性样本判断为阴性的概率;阴性预测值 (negative predictive value) 是当结果是阴性时真阴性的概率。灵敏度和特异度常用于描述定性方法的性能,而在临床中更关注阳性预测值和阴性

预测值。

对于定性和半定量检测项目,还可以评价实验室内部的一致性和实验室间的一致性 (类似定量检测项目的重复性和复现性)^[23],即内部一致性 (accordance) 和外部一致性 (concordance)。内部一致性指两个相同检测样本在同一个实验室相同条件下给出相同结果的可能性;外部一致性指两个相同的检测样本在不同情况下给出相同结果的可能性。若外部一致性小于内部一致性,则提示实验室间的变异可能更需要受到关注。

对于将定量结果转换为定性和半定量结果的情况,比如规定临界值 (threshold value) 或切值 (cut-off value),如新生儿遗传代谢病筛查中的苯丙氨酸 (Phe),在 EQA 计划中可以设定部分样本的浓度在临界值附近以考察参加实验室的检测能力,而在评价滴度时也可以进行类似的设定^[24]。

3 定量结果图形展示

3.1 单参数单样本 图形是 EQA 结果评价时非常有效的工具^[12],可用直方图和箱式图等。

直方图是首先将数据区间分成等宽的区间,然后对每个区间进行数据计数,之后按照条形图显示,高度即为结果数量。直方图能直观的给出结果分布情况,以及众数的可能位置等。直方图区间的宽度和数量等可以使用不同的算法来确定^[25]。

箱式图主要基于三个百分位数,即第 25,第 50 (中位数) 和第 75 百分位数。箱式图中的长方形上下边缘是第 25 和第 75 百分位数,延长线的上下限是除外所检出的离群值后的所有数据的范围。在延长线之外的数值是标记出来的离群值,即,在 $P_{25} - 1.5 \times (P_{75} - P_{25})$ 和 $P_{75} + 1.5 \times (P_{75} - P_{25})$ 两个界限之外的数值。箱式图可以直观地比较数据间的差别,在多个参数存在时是比较理想的 EQA 图形展示工具。

3.2 单参数多样本 多个样本时,信息的合并可以考虑使用散点图,将实验室结果和指定值共同显示于图中。使用线性回归可以从图中直观地观察偏移,还能帮助解释分析性变异和发现离群值等^[26]。

对实验室两个批号样本的结果联合做散点图称为尤登图 (Youden),在这里可以使用原始数据也可以使用转换后的数据^[12, 27]。可以在尤登图上标出不同方法的稳健置信范围 (椭圆形的范围)^[28],比如置信水平在 95% 和 99% 等。稳健置信椭圆范围可以为方法间的偏移、结果的互通性分析等提供信息。

4 多个样本相同参数的信息整合

对于参数间检测相互影响的情况有很多,比如血清中的电解质、白细胞分类计数、球蛋白和清蛋

白等,应该看做一个整体,也就是参数间并不是相互独立的。而对某实验室结果的评价,是基于多元变量多个参数的结果与每个参数指定值组成的中心之间的多元距离来确定的。

Z 比分数是标准差的标准化度量,更适用于不同参数的信息整合。可以将 Z 比分数相加和,采用:①直接加和 Z 比分数(简称“SZ”);②SZ 除以所包含数据数量的平方根,生成调整过的 Z 比分数求和,即 RSZ;③各数据绝对值的加和(SAZ),或者他们的平方和(SSZ)。计算求和前应进行离群值的判断并剔除。极端的 RSZ 表示偏移的存在,极端的 SAZ 表示不精密度过大。

如果不能确定数据集合是否服从或者近似服从正态分布,可以先进行数据转换并剔除离群值后再进行评价。如果之后仍不能确定是否服从正态分布,则可以使用非参数方法来进行实验室能力的评价。实验室结果与指定值的差距可以通过差值的绝对值来进行分级。之后计算每个回报结果与指定值差值绝对值的百分位数。通过不同批号样本百分位数的中位数结果进行评分,分数越低则实验室的表现越好,较高的分数提示实验室的结果与指定值相对差别较大^[29]。

某个实验室所检测样本的结果和各指定值可以通过回归方程进行结果合并和评价,自变量是指定值,因变量是实验室的回报结果。通过回归分析能获得长期偏移(long-term bias, LTB)、长期变异系数(long-term coefficient of variation, LCVa)^[30],以及测量不确定度(measurement uncertainty, MU)^[31]等统计量。

5 室间质评结果的回报与反馈

不同的 EQA 提供者对结果的处理和反馈方式一般遵循:①应保证参加实验室在每一轮次结果回报后尽快收到反馈;②整年度的总结性报告在当年各轮次结果回报后应尽快反馈给所有参加实验室。

EQA 样本的检测要与日常工作的检测过程相同。目前很多临床检验项目还没有参考方法来确定指定值。为了避免基质效应,样品一般不用大量混合全血或血清。而目前一般的 EQA 样本都不具有互通性,方法间的差异并不一定能反映样本检测时出现的差异^[32]。具有互通性的样品能够使 EQA 组织者获得更多的信息,比如方法间的一致性。没有互通性的样品一般都要进行分组评价。组内可以评价实验室是否参照规范程序进行检测,但不能评价实验室的准确度^[1]。而具有互通性的样品可以评价实验室的偏移、正确度和准确度等。

对于在 EQA 中结果不合格的实验室,应在结果回报中给出建议。该实验室也应该努力进行不合

格原因分析并做好纠正措施和记录^[33]。即使 EQA 的结果都在可接受范围以内,也可能存在问题,比如结果接近评价限,或合格结果都在指定值一侧等。

6 讨论与总结

对于定量数据,目前还没有一个公认的最优方案来对公议值进行估计。有 EQA 组织者尝试使用自身数据库进行回顾性分析,估计未知分布统计参数的不确定性,比较不同方法对指定值估计的差异和实验室评分的不同^[11, 16, 34]。单独分组的最小结果数量一直都有争论。有研究称组内最小数据量应 ≥ 20 ^[18]。也有研究显示组内数据 ≥ 6 就可以获得可靠的 Z 比分数估计值^[11]。目前国内更倾向于 ≥ 12 为分组标准。

涉及到临床解释性说明的 EQA 计划的最终目的都是希望改善病人的预后^[35]。有研究分析 EQA 长期数据后发现参加时间越久,检测能力和水平提高越明显,同时还使用了纵向混合效应模型(longitudinal mixed effects models),将结果绝对残差的自然对数(log, LN)带入模型并统计各因素的影响情况^[36]。EQA 计划应随时更新,比如 ZIKA 病毒或 COVID-19 病毒肆虐期间,包括他们的变异株,这些新项目需要进行质量控制时,应结合实际情况及时开展新的 EQA 计划,并定期进行效果评价^[37]。EQA 计划还能够对检验项目的室内质控水平^[38]、参考区间的应用情况^[39]、以及与行业标准等进行比较中发挥作用^[40],并可以通过不同类型的评价给与相应的结果回报或者结果汇总。

综上,EQA 反馈结果要起到教育和改进的作用。很多问题都能通过运行 EQA 发现,但它仅是质量管理体系中的一个方面。基于 EQA 回报而采取的改进措施都是滞后的。实验室应在检验全过程进行质量保证并持续改进,EQA 实际上只能起到类似后期随访的作用。

参考文献:

- [1] MILLER W G, JONES G R, HOROWITZ G L, et al. Proficiency testing/external quality assessment:current challenges and future directions[J]. Clinical Chemistry, 2011, 57(12): 1670-1680.
- [2] STAVELIN A, ALBE X, MEIJER P, et al. An overview of the European Organization for External Quality Assurance Providers in Laboratory Medicine (EQALM) [J]. Biochemia Medica, 2017, 27(1): 30-36.
- [3] COUCKE W, SOUMALI M R. Demystifying EQA statistics and reports[J].Biochemia Medica, 2017, 27(1): 37-48.
- [4] WANG Lunan, PAN Yang, ZHANG Kuo, et al. A 10-year human hepatitis B virus nucleic test external quality assessment in China: Continual improvement[J]. Clinica Chimica Acta, 2013, 425: 139-147.
- [5] ZHONG Kun, WEI Wang, HE Falin, et al. Neonatal

- screening external quality assessment in China, 2014[J]. *Journal of Medical Screening*, 2015, 22(4): 175-181.
- [6] ZHONG Kun, ZHAO Yan, XIAO Yalin, et al. 8-year review of laboratory performance on blood lead level external quality assessment surveys 2006~2013 in China: continual improvement[J]. *Accreditation and Quality Assurance*, 2015, 20(1): 25-28.
- [7] SARMAHO G F, BORGES P P, FRAGA I C, et al. Treatment of bimodality in proficiency test of pH in bioethanol matrix[J]. *Accreditation and Quality Assurance*, 2015, 20(3): 179-187.
- [8] MANCIN M, TOSON M, GRIMALDI M, et al. Application of bootstrap method to evaluate bimodal data: an example of food microbiology proficiency test for sulfite-reducing anaerobes[J]. *Accreditation and Quality Assurance*, 2015, 20(4): 255-266.
- [9] BARNETT V, LEWIS T. *Outliers in statistical data*[M]. 3rd ed. New York: Chichester, John Wiley & Sons, 1994.
- [10] DAVIES P L. Statistical evaluation of inter-laboratory tests[J]. *Fresenius' Zeitschrift für analytische Chemie*, 1988, 331: 513-519.
- [11] COUCKE W, CHINA B, DELATTRE I, et al. Comparison of different approaches to evaluate External Quality Assessment Data[J]. *Clinica Chimica Acta*, 2012, 413(5-6): 582-586.
- [12] International Organization for Standardization. ISO 13528: Statistical methods for use in proficiency testing by inter-laboratory comparisons [S]. Geneva:Switzerland, ISO 13528, 2006.
- [13] 中国合格评定国家认可委员会. 能力验证结果的统计处理和评价指南文件 [S]. 北京: 中国合格评定国家认可委员会 GL02, 2014.
- China National Accreditation Service for Conformity Assessment. Guidance on Statistic Treatment of Proficiency Testing Results and Performance Evaluation [S]. Beijing: CNAS GL02, 2014.
- [14] 钟堃, 王薇, 李少男, 等. 稳健统计在全国临床化学检验室间质量评价的应用[J]. *中国计量*, 2010, 15 (9): 46-48.
- ZHONG Kun, WANG Wei, Li Shaonan, et al. The application of robust statistics for National External Quality Assessment in clinical chemistry [J]. *China Metrology*, 2010, 15 (9): 46-48.
- [15] ZHONG Kun, WANG Wei, ZHANG Chuanbao, et al. The study of statistical methods for evaluating the comparability of routine chemistry analytes among 3 routine laboratory measurement systems in China[J]. *Springer Plus*, 2016, 5(1): 1724.
- [16] WONG S K. Evaluation of the use of consensus values in proficiency testing programmes[J]. *Accreditation and Quality Assurance*, 2005, 10(8): 409-414.
- [17] DUEWER D L. A comparison of location estimators for interlaboratory data contaminated with value and uncertainty outliers[J]. *Accreditation and Quality Assurance*, 2008, 13(4): 193-216.
- [18] THOLEN D W. Statistical treatment of proficiency testing data[J]. *Accreditation and Quality Assurance*, 1998, 3(9): 362-366.
- [19] COUCKE W, CHARLIER C, LAMBERT W, et al. Application of the characteristic function to evaluate and compare analytical variability in an external quality assessment scheme for serum ethanol[J]. *Clinical Chemistry*, 2015, 61(7): 948-954.
- [20] ROSARIO P, MARTÍNEZ J L, SILVÁN J M. Comparison of different statistical methods for evaluation of proficiency test data[J]. *Accreditation and Quality Assurance*, 2008, 13(9): 493-499.
- [21] DE BIÈVRE P. Fitness for purpose is different from a performance specification[J]. *Accreditation and Quality Assurance*, 2007, 12(10): 501.
- [22] SANDBERG S, FRASER C G, HORVATH A R, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine[J]. *Clinical Chemistry and Laboratory Medicine*, 2015, 53(6): 833-835.
- [23] LANGTON S D, CHEVENNEMENT R, NAGELKERKE N, et al. Analysing collaborative trials for qualitative microbiological methods: accordance and concordance[J]. *International Journal of Food Microbiology*, 2002, 79(3): 175-181.
- [24] DEOM A, EL AOUAD R, HEUCK C C, et al. Requirements and guidance for external quality assessment schemes for health laboratories[S]. World Health Organization, 1999.
- [25] SCOTT D W. On optimal and data-based histograms[J]. *Biometrika*, 1979, 66(3):605-610.
- [26] COUCKE W, BLERK M V, LIBEER J C, et al. A new statistical method for evaluating long-term analytical performance of laboratories applied to an external quality assessment scheme for flow cytometry. [J]. *Clinical Chemistry & Laboratory Medicine*, 2010, 48(5):645-650.
- [27] YODEN W J. Graphical diagnosis of inter-laboratory test results[J]. *Journal of Quality Technology*, 1959, 15(11):24-28.
- [28] SHIRONO K, IWASE K, OKAZAKI H, et al. A study on the utilization of the Youden plot to evaluate proficiency test results[J]. *Accreditation & Quality Assurance*, 2013, 18(3):161-174.
- [29] EHRMEYER S S, LAESSIG R H. Alternative statistical approach to evaluating interlaboratory performance[J]. *Clinical Chemistry*, 1985, 31(1): 106-108.
- [30] MEIJER P, DE MAAT M P, KLUFT C, et al. Long-term analytical performance of hemostasis field methods as assessed by evaluation of the results of an external quality assessment program for antithrombin[J]. *Clinical Chemistry*, 2002, 48(7): 1011-1015.
- [31] MATAR G, POGGI B, MELEY R, et al. Uncertainty in measurement for 43 biochemistry, immunoassay, and hemostasis routine analytes evaluated by a method using only external quality assessment data[J]. *Clinical Chemistry and Laboratory Medicine*, 2015, 53(11): 1725-1736.