

检测系统间方法学比对实验数据处理方法的探讨

苏武锦

(南宁中心血站,广西南宁 530003)

摘要:目的 讨论不同检测系统间方法学比对实验的数据处理方法。方法 采用罗氏生化检测系统作为比较的检测系统(X)、自建检测系统作为待评价的检测系统(Y),对新鲜血清样品进行丙氨酸转氨酶比对检测,实验数据应用配对 *t* 检验、相关分析、回归统计、Bland-Altman 法对 2 种检测系统间的系统误差进行评估,比较几种数据处理方法的优劣。结果 采用配对 *t* 检验进行 2 个检测系统的比较,得到 $t=3.841, P<0.05$,提示二者的检测差异有统计学意义;基于 EP9-A2 文件提供的方法得到回归方程: $Y=0.9587X+0.2452$,相关系数(r)=0.997,在 $X_c=40.0$ U/L 处估算的系统误差(SE)为 1.4 U/L,2 种检测系统的结果具有良好的一致性;Bland-Altman 差异图显示配对数据差值的均数(\bar{x})=-1.1 U/L,95%一致性界限为-5.0 U/L、2.8 U/L,在临床上可以接受。结论 不同检测系统间的比对应以其差异的大小为评估基础,同时采取多种方法从不同角度进行联合评价以避免采用一种方法评价的局限性。

关键词:因素分析,统计学; 方法学比对; 系统误差

DOI:10.3969/j.issn.1673-4130.2013.09.037

文献标识码:A

文章编号:1673-4130(2013)09-1138-03

实验室医学质量要求标准 ISO/15189^[1]对检验结果的溯源性和可比性提出了明确要求,并指出方法学比较试验(比对试验)是实现溯源的途径之一。美国临床和实验室标准化协会(Clinical and Laboratory Standards Institute, CLSI) EP9-A2 文件提供了对比实验方法,即采取 2 种检测系统对标本进行双份测定,然后将分析所得数据进行相关回归分析,计算 2 种检测方法之间的偏差。目前,国内学者采用 EP9-A2 文件提供的方法进行了不同检测分析系统之间结果对比的应用研究^[2-4];也有学者采用基于 *t* 检验及相关系数等统计学方法进行结果一致性的比对验证^[5-6]。国外有人采用 Bland-Altman 法评价不同检测系统的一致性^[7-8]。为探讨在比对不同检测系统的实验数据时,应用何种统计学方法更为合理,笔者分别应用配对 *t* 检验、相关分析、回归分析、Bland-Altman 法对 2 种不同检测系统间的比对实验数据进行了评估,现报道如下。

1 材料与方 法

1.1 主要仪器、试剂与样品 自建检测系统(日立 7080 全自动生化分析仪、日本和光丙氨酸转氨酶测定试剂盒、日本和光校准品组合)为待评价的检测系统(Y);罗氏生化检测系统(罗氏 Cobas P800 全自动生化分析仪、罗氏 Cobas 丙氨酸转氨酶测定试剂盒、罗氏 Cobas 校准品组合)为比较的检测系统(X)。检测标本为 50 份新鲜血清,其分析物浓度的分布覆盖高、中、低水平。

1.2 评估及数据处理方法 采用上述 2 种检测系统分别对 50 份新鲜血清标本进行双份测定,用配对 *t* 检验、相关分析、回归统计、Bland-Altman 法对二者实验数据的系统误差进行评估。

1.2.1 配对 *t* 检验和相关系数验证^[8-10] 引入统计学公式:

$$\bar{x} = \frac{\sum d}{n}, S_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}, t = \frac{\bar{d}}{S_d/\sqrt{n}}$$

上式中 d 为比较的检测系统(X)与待评价的检测系统(Y)测定数据的差值, \bar{d} 为差值样本的均数, S_d 为样本差值的标准差, n 为配对样本的对子数^[9]。当 $t \geq t_{0.05(v)}$,即 $P < 0.05$ 时,差异有统计学意义;当 $t \leq -t_{0.05(v)}$,即 $P \geq 0.05$ 时,差异无统计学意义,此时再进行线性评价,用散点图验证二者间的线性相关性。

1.2.2 基于 EP9-A2 文件提供的方法 (1)进行方法内离群点的检验;(2)比较的检测系统(X)测定范围的检验,如相关系数(r) ≥ 0.975 (或 $R^2 \geq 0.95$),则认为 X 取值范围合适,直线回归统计的斜率和截距可靠;(3)计算回归方程: $Y=bX+a$;(4)计算方法间的系统误差:将医学决定水平浓度(X_c)代入回归方程,计算待评价的检测系统(Y)与比较的检测系统(X)之间的系统误差(SE), $SE=|Y_c-X_c|=|(b-1)X_c+a|$;当 SE 小于 1988 年美国临床实验室改进修正案(Clinical Laboratory Improvement Amendments of 1988, CLIA'88)所规定允许误差的 1/2 时,提示 2 个检测系统的结果具有良好的一致性。

1.2.3 Bland-Altman 法^[7-8] 在二维坐标图上(图 1),用 X 轴表示 2 种方法测量结果的平均值,Y 轴表示 2 种方法测量值之间的差值;再取 $a=0.05(5\%)$,此时差值的 $(1-a)=95\%$ 的一致性界限为 $-1.96 SD, +1.96 SD$ 。图 2 中,上、下 2 条水平线代表 95%一致性界限的上、下限,中间实线代表差值的均线,中间虚线代表差值均数为 0($Y=0$)的情况。当代表均线的实线靠近虚线($Y=0$)时,表明 2 组测量数据的平均值差异较小,2 种测量方法的系统误差较小,95%的比对点都在一致性区间内,且在临床上可接受的,则认为这 2 种方法具有较好的一致性;如果其中一种方法测量结果的数值总是较高,就会得到所有点都高于或低于零线($Y=0$);如果差异图上的点散布如喇叭型(呈“<”,数值越大,越分散;或呈“>”数值越大,越集中),则表示变异不相等;若差异图上点的分布可以找到一条具有非零斜率的趋势线,这表明存在的比例误差。

1.3 统计学处理 用医学统计软件 MedCalc 11.4.2.0 作 Bland-Altman 图,其他数据用 SPSS18.0 软件及 Microsoft Office Excel 2003 进行统计学分析,2 个检测系统间的比较采用配对 *t* 检验、相关分析、回归法、Bland-Altman 法,以 $P < 0.05$ 为差异有统计学意义。

2 结 果

采用配对 *t* 检验进行 2 个检测系统的比较,得到 $t=3.841, P < 0.05$,提示二者的检测差异有统计学意义。基于 EP9-A2 文件提供的方法得到回归方程式: $Y=0.9587X+0.2452$,相关系数(r)=0.997,图 1 中虚线为理想状态($Y=X$)。在 $X_c=40.0$ U/L 时,估算的系统误差(SE)为 1.4 U/L,小于允许误差的 1/2(4.0 U/L),提示 2 个检测系统的检测结

果具有较好的一致性。采用 Bland-Altman 法, 配对数据差值的均数(\bar{x}) = -1.1 U/L, 均线的实线靠近虚线($Y=0$), 差值的标准差(SD) = 1.99 U/L, 95% 一致性界限为 -5.0 U/L、2.8 U/L。4.0% (2/50) 的点在 95% 一致性界限以外, 96.0% (48/50) 的点在一致性界限范围内, 待评价的检测系统(Y)与比较检测系统(X)测得的丙氨酸转氨酶的数值相比, 差值的绝对值最大为 4.0 U/L, 见图 2, 这种差异幅度在临床上可以接受, 因此, 2 种检测方法测量的结果具有较好的一致性。

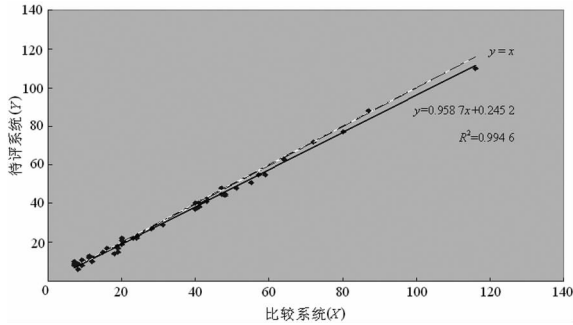


图 1 两种检测系统的结果比对

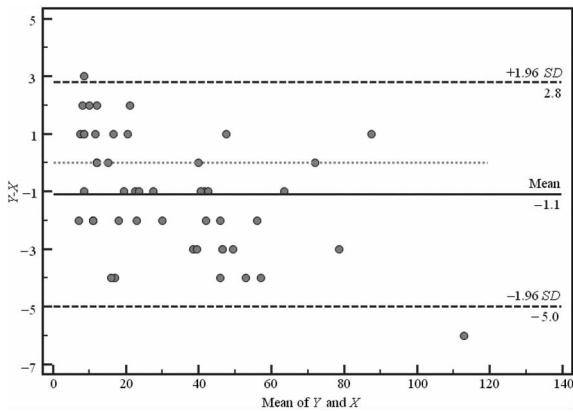


图 2 两种检测系统测量丙氨酸转氨酶的 Bland-Altman 图

3 讨论

临床实践中, 方法学比较方法有配对 t 检验、相关分析、回归方法、Bland-Altman 法等, 不同的分析方法各有其优、缺点。在方法比较实验中, 每份样品分别用 2 种检测系统检测, 得到的数据形成配对资料, 此时, 人们会想到配对 t 检验。如本实例, 计算得到 $t = 3.841, P < 0.05$, 提示存在系统误差, 但这并不能认为 2 种实验方法的吻合程度不可接受。在方法学比较中, 如果使用配对 t 检验来验证吻合程度, 假设比较的 2 个方法无差异, 即 2 个方法对每份标本检测结果的差为 0, 则配对 t 检验为:

$$t = \frac{|\bar{d}|}{S_d/\sqrt{n}} = \frac{|\bar{d}|}{S_d} \cdot \sqrt{n} = \frac{|\bar{X} - \bar{Y}|}{S_d} \quad (\text{相当于} \frac{|\text{系统误差}|}{\text{随机误差}})$$

该等式说明 t 检验数据受实验设计和系统精密度的影响, t 与 S_d 呈反比, 与 \sqrt{n} 成正比, 这在很大程度上使 t 值不可靠。

还需要指出的是 t 检验显示的差异无论是否显著, 与差异大小无关, 它反映的是 2 个检测系统检测结果差异的趋势性, 然而临床的实用性在于方法间的差异有多大而不是差异的趋势, 临床意义依赖于规定的允许误差。如观测误差小于允许误差, 方法性能可接受; 如观测误差大于允许误差, 方法性能不可接受。因此, 使用配对 t 检验, 应先计算所有成对数据差异的均值, 再计算差异的标准差, 对方法间的随机误差进行估计, 因为差异的标准差是方法间的离散度, 它取决于进行比较的 2 种

方法各自的不精密密度以及各种影响因素, 最后计算 t 值。正如上述等式表明, t 值本身相当于比较实验中系统误差和随机误差的比值, 主要用于确定偏倚是否有统计学上的显著性, 但不能用 t 检验结果作为临床意义上实验方法可接受的标准。

本实例中, 基于 EP9-A2 文件提供的方法, 笔者拟合回归方程: $Y = 0.9587X + 0.2452$, 并在 $X_c = 40.0$ U/L 时, 估算的系统误差 (SE) 远小于允许误差的 1/2, 这显示 2 种检测系统的结果具有良好的一致性。通过 Bland-Altman 法绘制 Bland-Altman 图, 可直观了解测量差值与测量均值的关系, 还清楚地显示了极端情况, 结合临床实际, 笔者认为这 2 种检测系统之间的系统误差可接受。若使用配对 t 检验, 显示 2 种检测系统的检测结果之间有显著性差异或 2 种检测系统之间无显著性差异, 但其平均差值仍较大时, 可借助 EP9-A2 文件提供的方法, 用回归分析来估计在决定性水平浓度下的系统误差是否在临床上可接受或考虑采用 Bland-Altman 法。

尽管 EP9-A2 文件提供的方法解决了如何评价测定同一样本的 2 种方法或仪器之间偏倚的一大难题, 但其应用过程必须谨慎。回归分析前, 首先, 进行方法内离群点的检验。其次, 以相关系数 (r) 估计比对数据时, 其分布范围要适当; 进行回归分析时, 斜率和截距的估计较可靠; 当相关系数较差, 应改善数据分布范围或选择其他统计方法, 冯仁丰^[10] 认为可将数据分成几组, 选用组均值和临床决定水平较一致的那组数据, 使用配对 t 检验中的差异均值 (偏倚) 进行分析。

对于 Bland-Altman 法的研究, 有人发现在模拟没有系统性比例偏差的数据中, Bland-Altman 差异图却呈现出有比例偏差, 认为用回归方法来分析有效性较适合; 也有人认为 Bland-Altman 差异图的问题值得再讨论, 使用回归方法来分析有效性值得考虑。

需要指出的是方法比较的研究中相关系数 (r) 只代表 2 个变量间关系的强度, 并不是 2 个变量的一致性。最好的相关系数 ($r = 1$) 只意味着一个变量数值的增加与另一变量数值的增加呈比例关系, 并不意味 2 个变量数值是一样的, 它仍可能存在系统误差。如一个变量数值高于另一变量数值 100 单位, 或单方面改变某一方法的度量单位 (如由 g/L 变为 mg/dL), 并不影响 2 个测量方法的相关性, 但这影响了 2 个测量方法的一致性。

综上所述, 在方法学比较中, 评价方法有多种, 各自有一定的局限性。在评价分析过程中, 应以系统间差异的大小为评估基础, 同时可以采取多种方法从不同角度进行联合评价, 以避免某一方法评价中局限性的影响, 使评价结论更为合理、客观。

参考文献

- [1] 魏昊, 丛玉隆. 医学实验室质量管理与认可指南 [M]. 北京: 中国计量出版社, 2004: 72-75.
- [2] 梅燕萍, 谭明娟, 张瑞生, 等. 两台血细胞分析仪的测定结果进行方法对比及偏倚评估 [J]. 国际检验医学杂志, 2012, 33(7): 858-859.
- [3] 李山, 黄鹏, 易珍, 等. 同一厂家不同型号血液分析仪检测结果的可比性研究 [J]. 国际检验医学杂志, 2009, 30(9): 833-834.
- [4] 王淑娟, 张敏, 宋予娟, 等. 不同检测系统间结果的可比性与临床可接受性研究 [J]. 国际检验医学杂志, 2011, 32(16): 1879-1880.
- [5] 刘兰民. Sysmex XE-2100 与 XT-1800i 全自动血分析仪的对比分析 [J]. 检验医学与临床, 2010, 7(16): 1747-1748.
- [6] 郝少丽, 丁晓旭, 王翠玲. 全自动血细胞分析仪 XE-2100 与 XS-1000i 的比较分析 [J]. 现代预防医学, 2009, 36(13): 2520-2521.

2527.

[7] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement[J]. Lancet, 1986, 1(8476):307-310.

[8] Ludbrook J. Confidence in Altman-Bland plots; a critical review of the method of differences[J]. Clin Exp Pharmacol Physiol, 2010, 37(2):143-149.

[9] 孙振球. 医学统计学[M]. 3 版. 北京: 人民卫生出版社, 2010: 38-39.

[10] 冯仁丰. 临床检验质量管理技术基础[M]. 上海: 上海科学技术文献出版社, 2007: 188-190.

(收稿日期: 2012-11-06)

• 质控与标规 •

6σ 质量管理方法在临床实验室制定校准周期中的应用

邓庆丰

(深圳宝安区中医院检验科, 广东深圳 518133)

摘要:目的 应用六西格玛(6σ)质量管理方法制定常规生化检验项目的校准周期。方法 收集该科 2010 年度临床检验项目室内质量控制及室间质量评价的数据,按允许总误差(TEa)标准,计算变异系数(CV)、偏倚、总分析性能 δ 值和不精密度 δ 值。总分析性能 δ ≥ 6(A 组)校准周期定为: > 30 d;总分析性能 δ < 6 且不精密度 δ ≥ 6(B 组)校准周期定为: 30 d;不精密度 δ < 6(C 组)在生化分析仪校准后,距校准时间 0、2、4、6、8、10、12、24 h 时分别测定 2 种浓度水平的质控血清各 1 次,其后每 24 h 测定 1 次,总计测定 30 d,计算累积不精密度 δ 值,以累积不精密度 δ < 6 为判断标准,确定被检生化项目的最长校准周期。严格按校准周期执行校准,1 年后,收集该科 2011 年度临床检验项目室内质量控制及室间质量评价的数据,2011 年校准周期在 30 d 以内所有项目的 CV、偏倚、TE 与 2010 年度进行比较。结果 该科共有 25 个生化检验项目,校准周期大于 30 d 的有 13 个,占 52%;校准周期在 30 d 内的有 12 个(含 1 个校准周期为 30 d),占 48%。2011 年校准周期在 30 d 以内所有项目的 CV、偏倚、TE 与 2010 年度进行比较,差异有统计学意义(P < 0.05)。结论 6σ 质量管理是临床实验室开展质量控制的一项有效管理模式,可用于校准周期的制定。

关键词: 质量控制; 校准; 六西格玛

DOI: 10. 3969/j. issn. 1673-4130. 2013. 09. 038

文献标识码: A

文章编号: 1673-4130(2013)09-1140-02

健全的室内质量控制系统是临床实验室质量管理工作的基础,是保证实验室检验结果准确、可靠的重要措施。而校准又是质量保证的前提,六西格玛(six sigma, 6σ)质量管理是近年来国际上迅速发展的一项以数据为基础,顾客为中心的先进质量管理模式^[1]。本科拟采用 6σ 质量管理方法来制定检验项目校准周期,对持续的质量改进帮助很大。

1 材料与方 法

1.1 主要仪器与试剂 主要仪器为日立 7180 型全自动生化分析仪,每半年厂家工程师进行一次仪器校准。主要试剂为 WAKO 生化试剂,(1)清洗液:抗菌无磷清洗剂(HITERGENT)、酸性清洗液(HICARRYNON)及碱性清洗液(HIALKALFD);(2)所有项目诊断试剂;(3)电解质参比液、内标液、稀释液;(4)所有项目定标液(有溯源)。质控品为美国 Bio-Rad 公司非定值生物物质质控物。

1.2 方 法

1.2.1 δ 值的计算及分组 收集本科 2010 年度临床检验项目室内质量控制及室间质量评价的数据,按照美国临床实验室改进修正法案允许总误差(allowed total errors, TEa)标准^[2],(1)计算变异系数(coefficient of variation, CV)和偏倚:临床检验项目的 CV 数据来源于本科 2010 年间室内质量控制数值,由于每个检验项目的质控有 2 个浓度水平,取 CV 值大者作为评估依据。偏倚数据来源于本室 2010 年间参加卫生部临床检验中心室间质量评价结果数值^[3];(2)计算:总分析性能 δ 值 = [TEa - 偏倚]/CV,不精密度 δ 值 = TEa/CV。根据 δ 值分 3

组,A 组:总分析性能 δ ≥ 6;B 组:总分析性能 δ < 6 且不精密度 δ ≥ 6;C 组:不精密度 δ < 6。

1.2.2 校准周期的制定 总分析性能 δ ≥ 6(A 组)校准周期定为: > 30 d;总分析性能 δ < 6 且不精密度 δ ≥ 6(B 组)校准周期定为: 30 d;不精密度 δ 值 < 6(C 组)在生化分析仪校准后,距校准时间 0、2、4、6、8、10、12、24 h 时分别测定 2 种浓度水平的质控血清各一次,其后每 24 h 测定 1 次,总计测定 30 d^[4-5],计算累积不精密度 δ 值,δ = TEa/累积 CV,以累积不精密度 δ < 6 为判断标准^[3,6],确定被检生化项目的最长校准周期。

1.2.3 执行校准周期后的误差比较 校准周期制定后,严格按校准周期执行校准,1 年后,收集本科 2011 年度临床检验项目室内质量控制及室间质量评价的数据,计算所有项目总误差(total error, TE): TE = 1. 96CV + 偏倚,计算 95% 允许误差限^[2,7]。2011 年校准周期在 30 d 以内所有项目的 CV、偏倚、TE 与 2010 年度进行比较,分别对 CV、偏倚、TE 的 2 年数据进行配对资料 t 检验。判断校准周期执行后质量是否有所改进。

1.3 统计学处理 δ 值及误差参数均由 Excel 2003 电子表格进行计算,采用 SPSS19.0 软件进行统计学分析,组间比较采用配对资料 t 检验,以 α = 0.05 为检验水准,以 P < 0.05 为差异有统计学意义。

2 结 果

分组后,确立各检测项目的校准周期,见表 1。本科共有 25 个生化检验项目,校准周期大于 30 d 的有 13 个,占 52%;