

诊断性检验中误差网格的建立与解释

曾 蓉¹, 王治国^{2△}

(1. 北京协和医学院研究生院 100730; 2. 卫生部北京医院卫生部临床检验中心 100730)

DOI:10.3969/j.issn.1673-4130.2011.14.069 文献标识码:B 文章编号:1673-4130(2011)14-1652-02

体外诊断结果在医学决策中的作用日益重要,因此,诊断性检验的质量受到更多的关注。虽然 Westgard 等^[1]表明诊断性检验中总分析误差很重要,但通常的做法是估计总误差。而且,即使按照 Bland 和 Altman^[2]的建议直接计算总误差,但对于这些误差可能对患者带来的损害的了解还是不够的。譬如在当今流行的六西格玛方法^[3]和测量不确定度的指南中,使用的医学误差界限只有 1 组^[4]。这意味着,超过这些医学误差界限并对患者有潜在伤害的值并没有受到限制。因此,需要应用误差网格来提供误差的全面信息。

1 基本概念

Clarke 等^[5]最早提出误差网格概念,并将其用于评估糖尿病患者所用的葡萄糖自我检测(SMBG)系统。但是由于其可以不经过中间区域便跨越 2 个区域,并且曲线不光滑,因此随后产生了 Parkes 共识误差网格^[6]。Parkes 共识误差网格被认为是 Clarke 误差网格的明显改进,并且是目前评价 SMBG 系统性能的行业标准。

误差网格是可以告知使用者试验方法中误差可能带来的临床后果的图。试验方法可为考虑的任何方法,可对比的方法通常是参考方法或可溯源至参考方法的方法。然而,可对比的方法可以是能提供有利于患者治疗价值的非参考方法。

的形状和位置在不同的检验方法中可能不同,但在 3 个区域的误差网格中,A 区和 C 区是绝对不能相接的。

2 误差网格与检验

2.1 准备误差网格 应结合临床试验结果与当前试验方法分析能力来准备误差网格。在确定 LER 区时,需要考虑与有害的临床结果相关的错误试验结果。可能存在与特定低和高水平试验结果相关的风险评估。在数据收集前,性能要求或可接受准则应该达成一致。

2.2 设定目标 每个区域的目标包括 2 个部分:区域中可允许误差的大小和区域中所允许结果的比例。在大部分应用中,性能要求的形式通常为“至少有目标百分数(T%)的测量点在 ATE 界线或 A 区”。典型情况下目标百分数很高,例如 T%为 90%、95%或 99%。在 LER 区中的结果百分数被设定为零或极小的比例。

2.3 测试误差网格 在测试误差网格中,需要考虑所采集的数据的代表性和样本量大小的问题。理想情况下,应该采用患者样本。数据必须要在测量仪器或方法被评估过的操作条件下才具有代表性。在数据收集阶段,确认关键变异的来源并充分理解。

3 结果百分数的定量与解释

3.1 置信区间 对误差网格各个区域中的百分数的计算是“点估计”。然而,为表达因样本变异导致的重复试验的不同值,可以采用如下公式计算置信区间:

$$P_L = \left\{ 1 + \frac{(n-x+1)F_{(1-\frac{\alpha}{2}; 2n-2x+2, 2x)}}{x} \right\}^{-1}$$
$$P_U = \left\{ 1 + \frac{n-x}{(x+1)F_{(1-\frac{\alpha}{2}; 2x+2, 2n-2x)}} \right\}^{-1}$$

其中 $F(y, r_1, r_2)$ 是指自由度为 r_1 和 r_2 的 F 分布的第 100 r 位的百分位数 x 为每个区域中观察的数量; n 为总的样本量; P_L 为置信区间下限, P_U 为置信区间上限。

当误差网格完成时,采用预先定义的性能要求来评价结果。如果结果不符合要求,则可以考虑增加样本量。此外,应该调查导致系统无法满足要求的原因。

3.2 解释 如果数据已经收集完,就可以得到是否达到目标的信息。C 区的目标达到而 A 区没有达到的情况比 A 区目标达到而 C 区没有达到更好; C 区目标没有达到而有人知道原因(如系统受到了试剂携带污染的影响)的情况比没有人知道 C 区目标失败的原因更为严重,因此,厂商或用户必须要处理已知的问题,对于未知的问题,应该进行调查;通过告知使用者这种危险情况来处理 C 区目标失败的特定原因;最后权衡医疗

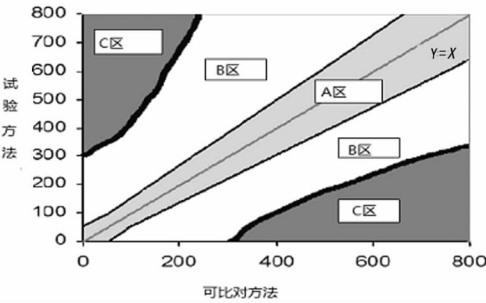


图 1 一般误差网格图

图 1 为典型的误差网格图。A 区,即 ATE 区,是包含了 $Y=X$ 线的最里面的区域,并且包含了大部分数据,其中的结果不可能对患者产生损害。C 区,即 LER 区,它是在最外面没有或仅包含几个数据的区域,其中的结果可以造成潜在的患者损害。当所划分的区域大于 3 个时,造成最大的潜在患者损害的区域用更高排序的字母来表示。B 区包含了 A 区和 C 区之间的结果,通常包含大于 5% 的数据,其中可能发生一些潜在的患者损害,但是比 C 区中的损害小,也可能完全没有损害。未报告的结果并非误差网格中的实际区域,其考虑的是试验方法不能获得结果时,可能引起患者损害的情况。尽管这些区域

△ 通讯作者, E-mail: zhiguo_w@hotmail.com.

的利弊,决定检验是否可接受。如果厂商已经进行了误差网格研究,那么这个研究的内容及其结果可以作为一个声明。

3.3 获得更多数据 计算误差网格的原始数据为每个误差网格区域提供了初步的结果百分数估计值,但是它们仅仅是整个可能的试验结果群中的一个样本。在初步的误差网格制作好之后,可通过重复比对的原始研究来确认数据的一致性。

4 误差网格在尿清蛋白检测中的应用实例

4.1 背景 尿微量清蛋白是慢性肾脏疾病(CKD)的早期标志,同时也是心血管疾病(CVD)的危险标志物^[7-8]。早期的诊断和预防,对糖尿病肾病很重要的。预防末期肾病的普通做法是筛查糖尿病患者中的尿微量清蛋白。

因为大多数国家规定在进行医学治疗之前要利用连续的样品对所有的尿清蛋白阳性结果(>40 mg/L)进行验证。因此患者的假阳性结果(试验值高于 40 mg/L 而可比对值低于 20 mg/L)对其没有很大的危害。而因为阴性结果(<40 mg/L)不需要重复检测,除非是糖尿病和高血压,因此这样假阴性结果(试验值低于 20 mg/L 而可比对方法的值高于 40 mg/L)会导致病情的延误。因此,利用误差网格评价试验方法的性能对患者安全十分重要。

4.2 构建误差网格 根据有关患者安全的类似讨论,可以建立误差网格^[9]。CM 表示可比对方法, TM 表示试验方法。

第 1 步,确定 $Y=X$ 线。

第 2 步,建立 ATE 区域。ATE 区域由生物学变异和临床目标共同决定。ATE 的标准如下:CM 值小于 33.3 mg/L 时, TM 值应在 CM 值 ± 10 mg/L 之内;CM 值大于 33.3 mg/L, TM 值应该在 CM 值的 $\pm 30\%$ 之内。结果在 ATE 内, ATE 的偏差小于 30% 或 10 mg/L, 这表明 CM 和 TM 导致的治疗决策无差别;结果在 ATE 外或 LER 外, 偏差大于 30%, 但是对患者而言不是危险的错误。

第 3 步,建立 LER 区域。当 TM 值落在 LER 区域内, 如果用这些值来做医疗决策的话, 潜在的患者损害可能发生。LER 区域定义如下:如果 CM 值小于 20 mg/L 及 TM 值大于 40 mg/L, 错误的结果可能导致错误的治疗, 但是对患者而言还不是危险的错误;如果 CM 值大于 40 mg/L 及 TM 值小于 20 mg/L, 当开始干预时未能检出错误的结果, 对患者而言虽然没有立即的危险但是长期效应将会很严重。

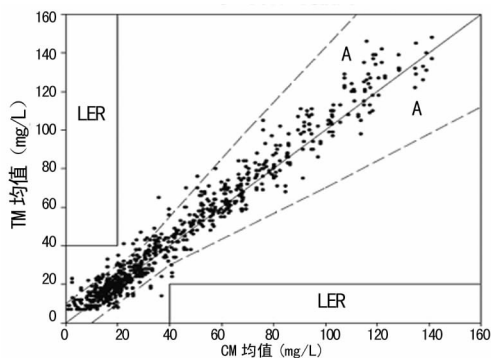


图 2 尿清蛋白的误差网格数据

4.3 数据填入与结果解释 接下来将方法比对数据填入误差网格中。数据来自在 5 个基层医疗中心和 1 个医院中开展的超过 18 d 的尿清蛋白检测的临床研究。所有的操作人员都是

有同等高中教育但是没有接受过正式的实验室训练的助理护士,他们没有受到过 TM 使用培训。总共有 424 份患者样本,重复测量 2 次,总共 848 个结果,见图 2。在试验方法空白检测限(即 7 mg/L)以下的观察值共有 49 个,而且这些值的 TM-CM 比对不能被清楚地计算。95% 的结果在 ATE 区域中,因此满足 ATE 区域中结果数大于 95% 的标准。在 LER 区域中没有结果,因此可以得到结论:采用这个试验系统不会使患者受到损害。

5 结 论

误差网格恰当地利用了所有的数据,通过将误差分级分区的方式来判断试验方法可能带给患者的损害^[7-10]。根据网格中不同区域的数据比例及其置信区间的值,以及建立误差网格时所设定的目标或者性能准则,检验人员可以对试验方法的性能做出更为全面的评估。因此,误差网格是改进实验室检验质量,减少对患者有损害的误差,从而切实保障患者安全的一种好的方法。

参考文献

- [1] Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation[J]. Clin Chem, 1974, 20(7): 825-833.
- [2] Bland JM, Altman DG. Statistical methods for assessing agreement between 2 methods of clinical measurement[J]. Lancet, 1986, 1(8476): 307-310.
- [3] 刘忠民, 高月亭, 肖洪广, 等. 6 σ 质量标准在临床生化检验室内质量控制中的应用研究[J]. 国际检验医学杂志, 2010, 31(3): 226-228.
- [4] 黄永富, 许文荣, 曹兴建. 全自动生化分析仪检测系统过程能力与不确定度的研究[J]. 国际检验医学杂志, 2010, 31(7): 650-652.
- [5] Clarke WL, Cox D, Gonder-Frederick LA, et al. Evaluating clinical accuracy of systems for self-monitoring of blood glucose[J]. Diabetes Care, 1987, 10(5): 622-628.
- [6] Parkes JL, Slatin SL, Pardo S, et al. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose[J]. Diabetes Care, 2000, 23(8): 1143-1148.
- [7] Hillege HL, Janssen WM, Bak AA, et al. Microalbuminuria is common, also in a nondiabetic, nonhypertensive population, and an independent indicator of cardiovascular risk factors and cardiovascular morbidity[J]. J Intern Med, 2001, 249(6): 519-526.
- [8] Kjeldsen SE, Westheim AS, Os I. Prevention of cardiovascular events and diabetes with angiotensin-receptor blockers in hypertension: Life, Scope, and Value[J]. Curr Hypertens Rep, 2005, 7(3): 155-157.
- [9] CLSI. Clinical and Laboratory Standards Institut. EP27-P How to construct and interpret an error grid for diagnostic assays: approved guideline[S]. Wayne, PA., 2009.
- [10] 阳辛, 张莉萍, 李武县, 等. 6 σ 质量管理理论在电化学发光法测定肿瘤标志物中的应用[J]. 重庆医学, 2010, 39(24): 3342-3343.

(收稿日期: 2011-01-24)