

骨密度测量的统计质量控制

方积乾 王显红

在骨质疏松症临床与科研中,骨密度(bone mineral density, BMD)作为一项重要的指标而被广泛应用。不管采用哪种方法,BMD测定值均在不同程度上受到操作人员、环境、仪器状况、测量方法、测量部位等诸多因素的影响,存在较大的随机误差和系统误差。在干预或临床试验中,BMD的改变量本来就不大,极易被误差掩盖,有必要进行测量方面的质量控制,包括异常偏差的预警和“正常”偏差的校正。笔者就有关统计学方法作一概述。

1. 精密度

精密度(precision)反映仪器测定结果的可重复性,是评价检测方法可重复能力的指标,在BMD测量中有着重要的地位^[1]。精密度分绝对精密度和相对精密度2大类^[2]。绝对精密度用标准差(standard deviation, SD)表示;相对精密度用变异系数(coefficient of variation, CV)表示;SD越大,绝对精密度越差;CV越大,相对精密度越差。绝对精密度提供了测量误差的重要信息,但它依赖于测量单位,不便于比较几种不同的测量仪器或几种不同的测定方法。在许多情况下,如疾病诊断或监测疾病过程随时间的变化,相对精密度更受关注。

根据测定的时间间隔和BMD随时间变化的情况,精密度可进一步划分为短期和长期精密度^[2,3]。短期精密度反映仪器检测重复性的优劣,用于描述骨骼状态测定值变化的限度;长期精密度用于评价仪器的稳定程度。

与短期精密度相比,长期精密度受到更多随机因素的影响,如仪器的漂移、病人骨密度的变化以及与时间有关的技术变化等。因此它能更好地监测仪器性能的改变,但临床应用中缺少灵活性。长期精密度的计算较复杂,需考虑个体参数与时间的关系,为简化运算,多数情况下可假设为线性关系,其描述指标为估计标准误(SEE)^[2]。

$$SEE_i = \sqrt{\frac{\sum_{j=1}^{n_i} (X_{ij} - \hat{X}_{ij})^2}{n_i - 1}} \quad (1)$$

其中, X_{ij} 表示第*i*个观察对象的第*j*次BMD测定值, $i=1, \dots, m, j=1, \dots, n_i, \hat{X}_{ij} = \hat{a}_i + \hat{b}_i t_{ij}$ 为在时点 t_{ij} 个体测量值的期望值。

对于1组*m*个研究对象,其绝对长期精密度(RMS SEE)为

$$RMS\ SEE = \sqrt{\frac{\sum_{i=1}^m (n_i - 2) SEE_i^2}{\sum_{i=1}^m (n_i - 2)}} \quad (2)$$

同时,可计算相对长期精密度,即纵向CV。纵向CV的计算很复杂,随测量时间和频率的不同而不同,而绝对精密度不会这样。因此,同类仪器间的比较,用绝对长期精密度RMS SEE更合适^[2]。

2. 质量控制图

精密度是评价仪器是否正常运行的重要指标,临床上安装仪器后,立即用精密度进行评价,之后也用精密度这一指标来对仪器设备进行监控^[4]。在理想的情况下,维护得好的设备的测量值应随机分散在参考值周围。此参考值为进行BMD测定时所用体模(phantom)的测量值,不随时间的改变而变化,即保持常量。利用体模观察到的变异可以反映人体测量的变异,并可评价仪器的运作正常与否。绘制质量控制图可以直观地观测误差趋势,对仪器的稳定性做出评价,便于及早发现异常并采取措施。

(1) Shewhart 质控图

绘制Shewhart质控图前需要2个参数:均数 μ 和标准差 σ ^[2]。均数可以由体模的理论值或在正常情况下测量到的首批25个观察值求平均得到。无论何时Shewhart质控图显示出一个失控信号,就要重新校准仪器,并以紧随其后的25个观察值的平均值作为新的均数 μ 。标准差因仪器和厂家而不同。例如,对DEXA骨密度仪而言,根据报道的体模长期精密度数据,通常假定Hologic体模的BMD变异系数为0.5%,而Lunar体模的BMD变异系数为

0.6%^[5]。因此,标准差就是 0.005 和 0.006 分别乘以 Hologic 和 Lunar 的均数。

以每次测得的体模 BMD 值为纵坐标,时间顺序为横坐标,在均数 μ 及 $\mu \pm \sigma, \mu \pm 2\sigma, \mu \pm 3\sigma$ 处共画 7 条横坐标的平行线,其中均数 μ 所对应的平行线为中心线,此图即为 Shewhart 质控图,见图 1。

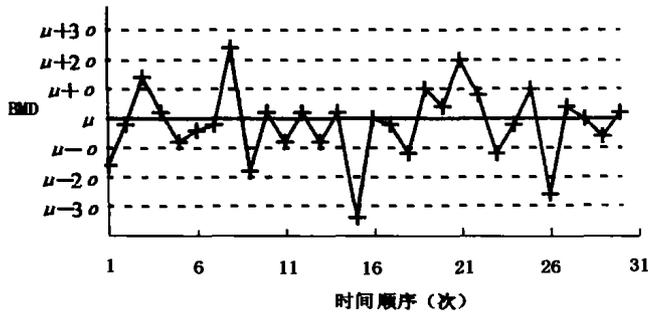


图1 体模 BMD 测定值随时间变化的 Shewhart 质控图

如果观察到体模测量值超出均数上下 3 倍标准差的范围,提示仪器失控,必须校准。这种方法易于应用,但对于虽小却有意义的变化,灵敏度不高^[6],可采用 1 套灵敏化规则来改进 Shewhart 质控图的灵敏度^[7],它包括以下 8 条规则:①有 1 个点与中心线距离超过 3 倍标准差;②在中心线的一侧连续有 9 个点;③连续 6 个点稳定地增加或减少;④连续 14 个点交替上下;⑤连续 3 个点中有 2 个与中心线距离超过 2 倍标准差;⑥连续 5 个点中有 4 个与中心线距离超过 1 倍标准差;⑦中心线一侧或两侧连续 15 个点与中心线距离都在一倍标准差范围;⑧中心线一侧或两侧连续 8 个点与中心线距离都超出一倍标准差范围。

这些灵敏化规则可以部分或全部应用。灵敏化规则增加了 Shewhart 图的灵敏度,但也增加了错误报警的次数。有专家提出了解决办法^[2],但同时指出其只对超出一倍标准差时才有效,可能会延误真正变点的确定。Shewhart 质控图对使用者的统计学知识要求不高,在临床研究中便于应用。

(2) 累积和质控图(CUSUM 图)

CUSUM 图是利用序贯分析的原理,以历次观测的累积结果判断仪器是否处于稳定状态。定义第 i 次质量控制测量值的单侧上限 $S_H(i)$ 和单侧下限 $S_L(i)$ 为

$$S_H(i) = \max\left[0, \frac{X_i - \mu}{\sigma} - \kappa + S_H(i-1)\right] \quad (3)$$

$$S_L(i) = \max\left[0, \frac{\mu - X_i}{\sigma} - \kappa + S_L(i-1)\right] \quad (4)$$

其中 μ 为均数, σ 为标准差, κ 为滤出无意义改变的

参数,通常取 0.5; $S_H(0)$ 和 $S_L(0)$ 的初始值为 0。当体模 BMD 值距离 μ 的值大于 κ , 累积的上界和会以超出 κ 值的量增加,如果偏差小于 κ , 累积和会以相应方式减少。当累积和小于 0 时,就忽略过去的数并设其为 0; 当累积和大于 5 时,足以说明数据偏离了均值。

以 $S(i)$ [含 $S_H(i)$ 和 $S_L(i)$] 为纵坐标,时间为横坐标作图,在 $S(i) = 5$ 处画一条横坐标的平行线(此为控制限),就得到了 CUSUM 图(图 2)。如果 $S_H(i)$ 或 $S_L(i)$ 大于 5, CUSUM 图就会报警,提示仪器失控,需要校正。

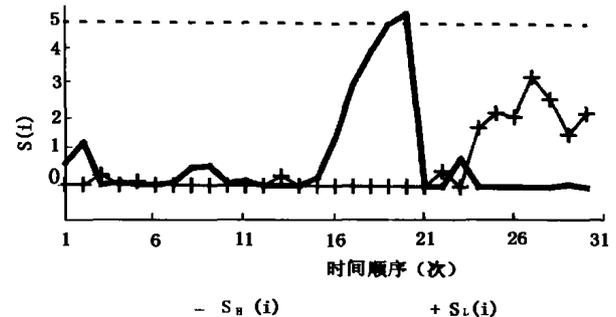


图2 体模 BMD 测定值随时间变化的 CUSUM 图

当 CUSUM 图显示出改变时,还能估计改变发生的时间和幅度,然后用估计出的改变幅度再建立新的参考值^[2](即均数和标准差)。

CUSUM 方法较好地结合了灵敏度、特异度以及变化时间和幅度的识别,因此适合在临床试验的质量控制中心使用。

(3) 其他图

上述方法及其他的质量控制图如移动平均图^[2]中,均假定在连续测量之间没有自相关,这对于骨密度仪的纵向质量控制来说几乎是不可能的。处理自相关的统计学方法是基于移除了自相关后的残差构造质控图,或使用指数加权的移动平均(EWMA)控制图^[6,8],用 EWMA 构成的质控图可以消除过多的误警问题,这是传统质控图无法克服的。

党容等^[9]在研究 DEXA 稳定性的统计学评价时发现, Shewhart 图存在 2 点不足:①在异常情形发生后才发出警报,虽然发现了问题,但未能提前预警,不能提供预报的有关信息;②不能排除漏报的发生。而 CUSUM 图缺乏对突然出现的个别极端值的有效灵敏控制。他们提出了若干补充准则,通过比较均数、标准差和检验自相关函数等方法对基线漂移、系统独立性进行检验,从而检测数据的稳定性。补充准则包括以下 3 条:①均数准则:($r-s$)个测定值的

均数水平与已知体模水平差异的 z 检验;②离散度准则: $(r-s)$ 个测定值的标准差与离散度控制阈值(体模均数的 0.5%)的 F 检验;③自相关准则: $(r-s)$ 个测定值的 1~5 阶自相关系数的统计学检验,检测序列是否为白噪声,即前后测量值之间是否相关,当时滞(lag)等于 1~5 的任一阶自相关系数有统计学意义($\alpha = 0.05$)时,认为存在自相关。

当均数准则和离散度准则中任一个假设检验呈阳性($P < 0.05$)时,则示警;当只有自相关函数的假设检验呈阳性时,则预警。

时滞等于 k 的样本自相关系数 γ_k 为

$$\gamma_k = \frac{\sum_{i=k+1}^n (y_{i-k} - \bar{y})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$(r-s)$ 段数据的个数由时间序列控制窗口 WINDOW 决定。WINDOW 宽度反映时间序列的记性,即各种因素(仪器状况、操作程序、环境等)对测量值的影响效果的持续时间。如果持续时间较长,相应的 WINDOW 宽度也要定得大一些。党容等^[9]在研究 DEXA 体模数据时,通过比较不同 WINDOW 宽度对异常数据的报警、预警情况,建议 WINDOW 的宽度取为 15。

随着 WINDOW 的依次滑动,计算各时间点对应的 WINDOW 内数据的均数、标准差和自相关系数并按补充准则进行假设检验,即可对数据的异常波动与否做出判断,从而对仪器的稳定性做出评价。

党容等^[9]通过 Shewhart 质控图灵敏度准则和补充准则对模拟数据判别的灵敏度和特异度的比较以及关联性分析,认为补充准则有很好的补充作用,两者结合使用,更具有普遍适用性。

3. 一致性评价

在骨质疏松症的临床研究中,不同研究中心不是使用同一台骨密度仪,为了结果的一致,必须保证多台仪器的 BMD 测量值的等价性或一致性。在试验过程中,可能会更新仪器,试验的有效性评价是考察相对于基准水平 BMD 的改变,因此也必须保证新旧仪器测得的 BMD 值的一致性。

设个体 i 在两台骨密度仪上的 BMD 测量值分别为 $Y_{1,i}, Y_{2,i}, i = 1, 2, \dots, n$, Y_1 和 Y_2 分别服从正态分布 $N(\mu_{Y_1}, \sigma_{Y_1})$ 和 $N(\mu_{Y_2}, \sigma_{Y_2})$, Y_1 和 Y_2 的相关系数为 ρ 。令 $D = Y_1 - Y_2, A = (Y_1 + Y_2)/2$, 进行 D 关于 A 的回归分析,得到回归系数的估计值 $\hat{\alpha}$ 和 $\hat{\beta}$ 。

容易证明:如果 $D = x + \beta A + \epsilon$, 则

$$\beta = Cov(D, A) / \sigma_A^2$$

$$= 0.5(\sigma_{Y_1}^2 - \sigma_{Y_2}^2) / (\sigma_{Y_1}^2 - \sigma_{Y_2}^2 + 2\rho\sigma_{Y_1}\sigma_{Y_2}) \quad (6)$$

$$\alpha = (\mu_{Y_1} - \mu_{Y_2}) - \frac{\mu_{Y_1} + \mu_{Y_2}}{2}\beta \quad (7)$$

因此, $\alpha = \beta = 0$ 就意味着 $\mu_{Y_1} = \mu_{Y_2}$ 和 $\sigma_{Y_1} = \sigma_{Y_2}$ ^[10], 即 2 台骨密度仪所测 BMD 具有相同的分布参数,即具有一致性。

可以用 Bradley-blachwood 方法同时检验假设 $H_0: \alpha = \beta = 0$ ^[11], 其检验统计量为

$$F = (n-2) \left[\sum_{i=1}^n D_i^2 - \sum_{i=1}^n (D_i - \hat{\alpha} - \hat{\beta}A_i)^2 \right] / \left[2 \sum_{i=1}^n (D_i - \hat{\alpha} - \hat{\beta}A_i)^2 \right] \quad (8)$$

H_0 成立时, F 服从 $F(2, n-2)$ 分布。

4. 校准和标准化

在临床或科研期间,常采用质量控制图来监控仪器是否在要求的精密度下运行,以及不同仪器之间是否保持一致。如果仪器设备不具有一致性,就必须选择某一标准对测定结果进行校准,即进行标准化。即使单台仪器,如果测量所得体模 BMD 值的均数与理论值的差别有临床意义,也同样需要校准。

国际骨测量标准化委员会(ICSBM)和国际双光子 X 线骨密度仪标准化委员会(IDSC)推荐了标准化校准公式,并在此基础上进一步完善^[12,13],将不同厂家仪器所测 BMD 值转化成标准化 BMD 值($sBMD$)。从而使不同厂家仪器的 BMD 测量结果具有可比性。但这种校准方式是否真正能达到使不同仪器测量结果等价的目的还有待研究,而且国外的公式不能照搬。要产生适合国情的校准公式,需要采用多种体模;每一种体模同时用几种仪器测定;最后,进行简单回归。陆盈等^[2]综合了在不同情况下(如单台仪器、多台仪器)对仪器结果进行校准的模型,为仪器的标准化提供了必要的工具。

5. 小结

在 BMD 测量的质量控制中,质控图的应用是一个重要的部分,不同质控图各有优劣。新近提出的补充准则利用时间序列分析方法对质控数据的自相关性进行了处理,弥补了传统质控图的不足,但同时也降低了报警的特异度,有必要进一步探索更有效的方法。对仪器进行标准化的方法也需要进一步完善,从而使之更符合具体情况。

同时, BMD 测量的质量控制不仅仅是一个统计学问题,它需要多方协作、共同把关。当用统计学方法发现仪器的精密度、稳定性不理想时,需要厂家、

(下转第 210 页)

参 考 文 献

- Pierson RN, Wang J, Thornton JC, et al. Bone mineral and body fat measurements by two absorptiometry systems comparisons with neutron activation analysis. *Calcif Tissue Int*, 1995, 56:93.
- Mazess RB, Trempe JA, Bisek JP, et al. Calibration of dual-energy X-ray absorptiometry for bone density. *J Bone Miner Res*, 1991, 6:799.
- Morita R, Orimo H, Hamamoto I, et al. Some problems of dual energy X-ray absorptiometry in clinical use. *Osteoporosis Int*, 1993, 3(Suppl): 87-90.
- 杨定焯, 尚家芸, 杨惠. 骨密度测量的质量控制及重要性. *中国骨质疏松杂志*, 2002, 8:365-366.
- Cawte SA, Pearson D, Green DJ, et al. Cross-calibration precision and patient dose measurements in preparation for clinical trials using dual energy X-ray absorptiometry of the lumbar spine. *Br J Radiol*, 1999, 72: 354-362.
- Laskey MA, Flaxman ME, Barber RW, et al. Comparative performance *in vitro* and *in vivo* of Lunar DPX and Hologic QDR-1000 dual energy X-ray absorptiometer. *Br J Radiol*, 1991, 64:1023-1029.
- Rosvold Berntsen GK, Fonnebo V, Tollan A, et al. The tromso study: determinants of precision in bone densitometry. *J Clin Epidemiol*, 2000, 53:1104-1112.
- 伍贤平, 廖二元, 陆泽元, 等. 双能 X 线吸收法测定大鼠骨量的评价及去卵巢骨丢失敏感区的选择. *中华内分泌代谢杂志*, 2000, 16:212-215.
- Kuiper JW, Kuijk VC, Grashius JL, et al. Accuracy and influence of marrow fat on quantitative CT and Dual-Energy X-ray Absorptiometry measurements of the femoral neck *in vitro*. *Osteoporosis Int*, 1996, 6:25-30.
- Bolotin HH, Sievanen H, Grashius JL, et al. Inaccuracies inherent in patient-specific DEXA bone mineral density measurements: Comprehensive phantom-based evaluation. *J Bone Miner Res*, 2001, 16:417-426.
- Sievanen H, Kannus PJ, Rvinen M. Precision of measurement by dual-energy X-ray absorptiometry of bone mineral density and content in rat hindlimb *in vitro*. *J Bone Miner Res*, 1994, 9:473-478.
- Gundry CR, Miller CW, Ramos E, et al. Dual-energy radiographic absorptiometry of the lumbar spine: clinical experience with two different systems. *Radiology*, 1990, 174:539-541.
- Mazess RB, Collick B, Trempe J, et al. Performance evaluation of a dual-energy X-ray bone densitometer. *Calcif Tissue Int*, 1989, 44:228-232.
- Tohill P, Fenner JA, Reid DM, et al. Comparisons between three dual-energy X-ray absorptiometers used for measuring spine and femur. *The British Journal of Radiology*, 1995, 68:621-629.
- Bouyoucef SE, Cullum ID, EH PJ, et al. Cross-calibration of a fan-beam X-ray densitometer with a pencil-beam system. *Br J Radiol*, 1996, 69:522-531.
- Laval-Jeantet AM, Roger B, Bouysse IS, et al. Influence of vertebral fat content on quantitative CT density. *Radiology*, 1986, 159:463-466.
- Crawley EO, Evans WD, Owen GM, et al. A theoretical analysis of the accuracy of single energy CT bone mineral measurements. *Phy Med Biol*, 1988, 33:1113-1127.
- Genant HK, Grampps, Gluer CC, et al. Universal standardization for dual X-ray absorptiometry: patient and phantom cross-calibration result. *J Bone Miner Res*, 1994, 9:1503-1511.
- Shepherd JA, Cheng XG, Lu Y, et al. Universal standardization of forearm bone densitometry. *J Bone Miner Res*, 2002, 17:734-745.
- Gluer CC. Monitoring skeletal changes by radiological techniques. *J Bone Miner Res*, 1999, 14:1952-1962.

(收稿日期:2003-04-10)