

## • 综述 •

## 以加权最小二乘法建立生物分析标准曲线的若干问题

钟大放

(沈阳药科大学药物分析研究室 沈阳 110015)

根据国际上有关生物分析方法确证的指导原则, 要求对不同浓度生物样品测量结果的相对误差不得超过 15% 或 20%。由于生物分析所测定的浓度范围宽, 若采用普通最小二乘法建立标准曲线, 在低浓度区域很难满足上述要求。本文通过总结作者多年从事生物分析工作的经验, 讨论了以加权最小二乘法建立标准曲线的优越性, 介绍了该法的原理、计算公式及应用实例。

在进行生物利用度、生物等效性及药物动力学研究中, 要对生物样品中的药物及其代谢产物进行定量分析, 一般采用标准曲线法进行数据处理。通过对系列浓度的标准生物样品进行测定, 得到浓度 ( $X_i$ ) 与响应 ( $Y_i$ ) 的对应数据, 再进行回归运算, 得到标准曲线方程 (在色谱分析方法中, 通常是线性方程)。然而, 生物样品具有分析物含量变化大的特点, 需检测的浓度范围宽, 若使用了不恰当的回归运算方法, 会得不到最佳的标准曲线, 甚至导致错误的计算结果。为解决这一问题, 本文讨论加权最小二乘法在建立生物分析方法标准曲线 (线性方程) 中的应用。

## 1 生物分析方法确证的有关国际规范

由于生物分析方法在评价与解释生物利用度、生物等效性及药物动力学数据中的重要作用, 对方法的确证 (validation) 已引起广泛关注<sup>[1~5]</sup>。1990 年 12 月, 几十个国家的专家学者在美国举行专门会议, 讨论制定了进行生物利用度、生物等效性及药物动力学研究时, 生物分析方法确证的指导原则<sup>[5,6]</sup>, 并在 6 种国际性学术期刊发表, 成为目前公认规范 (以下简称“规范”)。在规范中, 要求标准曲线要覆盖未知样品的整个浓度范围, 应用 5~8 个浓度点确定 (不含零点)。

规范提出, 应采用质量控制 (QC) 样品评价分析方法的准确性、精密度及测试结果的有效性。QC 样品取 3 个浓度, 分别在最低、最高检测限附近和待测浓度范围中部。在分析方法建立阶段和未知样品分析阶段, 一般要求 QC 样品的测量值相对误差分别在理论值的  $\pm 15\%$  或  $\pm 20\%$  范围内。

需要特别指出的是, 这里规定的是相对误差限。由于生物分析所测定的浓度范围宽, 故在不同浓度区域测量值的相对误差与绝对误差远远不成比例。例如, 在样品浓度为 1000 ng/ml 时, 测量值绝对误差 50 ng/ml 仅使其相对误差达到 5%, 而在样品浓度为 10 ng/ml 时, 同样大的测量值绝对误差则使其相对误差高达 500%。这提示我们对低浓度区域的 QC 样品的测量结果应给予特殊的关注。除了实验方法本身之外, 因为不同的回归运算方法对各浓度点相对误差的影响也很大, 所以能否选用恰当的计算方法以获得最佳的标准曲线, 常常会决定生物分析结果的质量。

## 2 通常用于建立标准曲线的几种回归方法的局限性

## 2.1 普通最小二乘法

普通最小二乘法<sup>[7]</sup>可求算  $Y=a+bX$  形式的回归直线方程参数 (截距  $a$  和斜率  $b$ )。由于许多类型的计算器上有此功能键, 故应用较广。对于色谱分析方法,  $Y$  一般为分析物与内标物的峰高 (或峰面积) 比值,  $X$  为样品中分析物浓度。这种回归运算的目标是使得观测值 ( $Y_i$ ) 对回归直线上对应估计量 ( $Y_i'$ ) 偏差的平方和最小, 即达到  $\min \sum (Y_i - Y_i')^2$ \*。对于生物分析方法来说, 这意味着标准曲线上每个浓度点的绝对误差具有同等的重要性。然而经验和理论都

\* 在本文中, 以简化的算符  $\sum_{i=1}^n$  代表

表明,生物样品测试的绝对误差会随浓度的增加而增加,在需检测的浓度范围(最高检测限与最低检测限浓度之比)为几十倍甚至几百倍时,使用普通最小二乘法进行回归运算,获得的标准曲线必然导致在低浓度区域测量值的相对误差甚大,很难满足规范的要求。

我们曾采用 HPLC 方法测定兔血浆样品中的某

药物的浓度。表 1、2 列举了一个典型的分析批 (Run) 标准曲线与 QC 样品的分析结果。从结果的相对误差容易看出,简单地使用未加权的最小二乘法,将导致低浓度区域分析结果的准确性明显下降。这一事实表明,为了满足规范的各项要求,保证分析结果的质量,必须采用其他回归方法建立生物分析标准曲线。

表 1 标准曲线 (根据 6 个样品的浓度与响应数据, 分别以 2 种方法求线性回归方程, 再将各样品数据代入这些方程, 计算测得浓度与相对误差)

加入浓度 (ng/ml)	分析物/内标物 色谱峰高比	普通最小二乘法结果		权重	加权最小二乘法结果	
		测得浓度 (ng/ml)	相对误差 (%)		测得浓度 (ng/ml)	相对误差 (%)
30.0	0.0616	42.9	43.1	11111	30.3	0.9
60.0	0.1152	71.5	19.1	2778	59.8	-0.3
150.0	0.2684	153.1	2.1	444	144.3	-3.8
600.0	1.091	591.4	-1.4	28	598.1	-0.3
1500.0	2.721	1460	-2.7	4	1497	-0.2
3000.0	5.648	3019	0.6	1	3112	3.7
截距 a		-0.0190			0.0067	
斜率 b		0.0019			0.0018	
相关系数 r		0.9998			0.9996	

表 2 QC 样品 (低、中、高 3 个浓度点, 每点 2 个样品, 由上述标准曲线分别计算测得浓度与相对误差)

加入浓度 (ng/ml)	分析物/内标物 色谱峰高比	普通最小二乘法结果		加权最小二乘法结果	
		测得浓度 (ng/ml)	相对误差 (%)	测得浓度 (ng/ml)	相对误差 (%)
60.0	0.1111	69.3	15.5	57.6	-4.1
60.0	0.1224	75.3	25.5	63.8	6.3
600.0	1.065	577.5	-3.7	583.7	-2.7
600.0	1.100	596.0	-0.7	602.8	0.5
3000.0	5.601	2994	-0.2	3086	2.9
3000.0	5.498	2939	-2.0	3028	0.9

2.2 两条回归直线法 对普通最小二乘法的一种改良措施是将测量浓度范围分为高、低 2 个区间, 每个区间分别用普通最小二乘法求得一条回归直线, 合起来作为标准曲线。这样, 由于每个区间浓度范围变窄, 测量值绝对误差的变化也相应减小, 部分地弥补了普通最小二乘法的上述缺陷。然而, 这一方法存在一些缺点: 按规范要求, 每条标准曲线至少应包括 5 个浓度点, 这样, 两条回归直线共将包括至少 8 个浓度点, 其中第 4 点和第 5 点之间的区段为 2 个区间共有, 因此, 需要计算 2 条回归直线的交点, 在进行未知样品

测定时, 需逐个考虑应用哪条回归直线计算浓度。这样, 在建立标准曲线时既增加了样品分析工作量 (多了 3 个浓度点), 计算工作也比较繁琐。

2.3 多项式拟合的回归曲线 用于生物样品测定的色谱分析方法, 其浓度与响应的函数关系通常呈线性, 仅在少数情形下, 在高浓度区域或极低浓度区域会发生对直线的偏离。因而有人采用多项式拟合曲线的回归方法, 一般是增设二次项, 即采用  $Y=a+bX+cX^2$  的形式。这在响应函数不呈线性时是可取的, 但若一般地以这种形式的标准曲线来代替普通最小二乘

法求算的回归直线, 则仍然不适宜。其原因在于, 一方面对于不呈线性关系的标准曲线, 规范要求测定更多 ( $\geq 8$ ) 的浓度点以确定浓度与响应的关系, 测试与计算的工作量都要增加; 另一方面, 如果这种回归运算仍旧是以  $\min \sum (Y_i - Y_i')^2$  为目标, 则同样不能很好地解决低浓度区域分析结果相对偏差畸大的问题。显然, 根据生物分析方法的特殊性, 回归运算只有以测量值相对偏差的平方和最小为目标, 即达到  $\min \sum [(Y_i - Y_i') / Y_i']^2$ , 才有可能获得最佳的标准曲线。

### 3 关于加权最小二乘法

#### 3.1 原理

近年来, 分析方法委员会 (Analytical Methods Committee) 推荐使用加权回归方程检查标准曲线的线性<sup>[7]</sup>。与普通最小二乘法不同的是, 加权最小二乘法 (weighted least squares) 在回归计算时增加了 1 个权重因子  $W_i$ , 即通过达到  $\min \sum W_i (Y_i - Y_i')^2$ , 来求算回归直线的斜率和截距。计算时, 一般使权重与绝对误差成反比, 亦即将大的权重赋予绝对误差小的点, 而小的权重赋予绝对误差大的点。把以这种方法求算的回归直线 (或曲线) 作为生物分析标准曲线, 可使生物样品的测定结果与理论值的相对偏差在不同的浓度区间内比较均衡, 易于满足规范的要求, 从一个方面保证了生物利用度、生物等效性与药物动力学研究数据的可靠性。

在理论上, 普通最小二乘法是加权最小二乘法的 1 个特例, 即每个浓度点的权重因子均相等。一般说来, 与加权回归直线相比, 未加权的回归直线  $r$  值更接近于 1, 在高浓度区域的测量结果相对误差略小, 但在低浓度区域的测量结果相对误差增大极为显著。显然, 不能简单地以  $r$  值的高低来评价这 2 种不同线性回归方法所获得的标准曲线的优劣。

#### 3.2 回归直线方程参数的计算公式 加权最小二乘法

回归直线参数方程的计算比未加权的要复杂。以下提供的是回归方程  $Y = a + bX$  中截距  $a$ 、斜率  $b$  及相关系数  $r$  的计算公式。公式的导出参见有关文献<sup>[7,8]</sup>。

$$a = \frac{(\sum W_i X_i^2) (\sum W_i Y_i) - (\sum W_i X_i) (\sum W_i X_i Y_i)}{(\sum W_i) (\sum W_i X_i^2) - (\sum W_i X_i)^2} \quad (1)$$

$$b = \frac{(\sum W_i) (\sum W_i X_i Y_i) - (\sum W_i X_i) (\sum W_i Y_i)}{(\sum W_i) (\sum W_i X_i^2) - (\sum W_i X_i)^2} \quad (2)$$

$$r = \sqrt{\frac{(\sum W_i (a + bX_i - \sum W_i Y_i / \sum W_i)^2)}{\sum W_i (Y_i - \sum W_i Y_i / \sum W_i)^2}} \quad (3)$$

实际应用时可编制 BASIC 程序或使用 EXCEL, NCSS 等计算软件, 在微机上计算。

#### 3.3 权重因子的选择

对于权重因子的选择依据是: 它应使各测定点具有适当的权重, 由此算得的标准曲线应尽可能使各浓度点测量值的相对误差都符合规范的要求。由式 (1) ~ (3) 可知, 各个  $W_i$  值乘以同一个常数  $K$  后, 对  $a$ 、 $b$  和  $r$  的值并无影响, 所以关键是确定各个  $W_i$  值之间的相对比例关系。在通常采用的方法<sup>[7,8]</sup>中, 是令  $W_i$  与测量值  $Y_i$  的方差成反比, 即

$$W_i = K / S_i^2 \quad (4)$$

这里,  $S_i$  的值可由适当的模型表达式求算, 或者通过对每个  $Y_i$  值的平行测量求得。但是, 我们在生物分析实际工作中采用最多的是

$$\text{令 } W_i = K / X_i^2 \quad (5)$$

$$\text{或 } W_i = K / Y_i^2 \quad (6)$$

我们的目标是通过选择权重因子使  $\min \sum [(Y_i - Y_i') / Y_i']^2$ , 由于  $Y_i'$  值预先未知, 这里的假定条件是  $Y_i'$  值与  $X_i$  值或  $Y_i$  值成正比。这样也非常便于用电子计算机处理日常分析数据。按式 (5) 和式 (6) 算出的结果往往非常接近, 可根据习惯选定 1 种。但在某些情形下, 按式 (5) 或式 (6) 确定权重因子显得对标准曲线上低浓度点加权过重, 从而导致高浓度点的测量值准确性损失过多。这时, 可适当降低低浓度点的权重, 如将最低浓度点的权重因子降至次低浓度点的水平。还可按式 (7) 或式 (8) 确定权重因子:

$$\text{令 } W_i = K / X_i \quad (7)$$

$$\text{或 } W_i = K / Y_i \quad (8)$$

总之, 权重因子有多种模式, 可根据每种分析方法的测量结果作出选择或调整, 然后确定下来。

#### 3.4 在生物分析中的应用

对表 1 中的数据以加权最小二乘法计算标准曲线, 采用式 (5) 给出的权重因子, 获得的结果比普通最小二乘法好得多 (见表 1、2 中相对误差的对比)。这是非常典型的情形。

作者多年来在国内外进行的近 30 项关于制剂的生物利用度、生物等效性及药物动力学研究中, 对于包括紫外、荧光、电化学检测的 HPLC 法以及电子捕获、质谱检测的 GC 法等诸多生物分析方法, 多采用加权最小二乘法求算单一的直线方程作为标准曲线, 既包括在方法建立阶段每个浓度点多重样品的情况, 也包括在未知样品分析阶段每个浓度点单一样品的情况。仅在测定浓度范围特宽 (如上下限比值在 100 倍以上) 的少数情形下, 采用不同浓度区间 2 条回归直线 (均用加权最小二乘法求算) 组合的方法。权重因子多在采用  $W_i = K / X_i^2$  时 QC 样品的测量准确性最

佳,即各浓度点测量值的相对误差均较小。在约10%的分析方法中,权重因子采用 $W_i=K/X_i$ 时最佳。

综上所述,对于保证生物分析数据的质量,采用何种算法建立标准曲线至为重要。本文介绍的加权最小二乘法针对生物分析方法测定浓度范围宽的特点,应用了适当的权重因子,对保证每个分析批的QC样品(尤其是低浓度的样品)测量准确性满足规范要求,起到了重要作用。至于普通最小二乘法,它仅适用于那些测定浓度范围窄的分析方法,如制剂含量或含量均匀度分析等,在建立生物分析标准曲线时应避免使用。

感谢本校研究生李雪庆为本文整理资料并提出建设性意见。

## 参考文献

- 1 Karnes H T, et al. Validation of bioanalytical methods. *Pharm Res*, 1991, 8: 421
- 2 Carr G P, et al. A practical approach to method validation

- in pharmaceutical analysis. *J Pharm Biomed Anal*, 1990, 8: 613
- 3 Causey A G, et al. Evaluation of criteria for the acceptance of bioanalytical data. *J Pharm Biomed Anal*, 1990, 8: 625
- 4 Buick A R, et al. Method validation in the bioanalytical laboratory. *J Pharm Biomed Anal*, 1990, 8: 629
- 5 Shah V P, et al. Analytical methods validation; bioavailability, bioequivalence, and pharmacokinetic studies. *J Pharm Sci*, 1992, 81: 30
- 6 萧 参等. 生物药剂分析方法的认证. *中国药理学杂志*, 1993, 28: 425
- 7 Miller J N. Basic statistical methods for analytical chemistry. Part 2. Calibration and regression methods. *Analyst*, 1991, 116: 3
- 8 许 禄. 化学计量学方法. 第一版. 北京: 科学出版社, 1995. 33

(本文于1995年10月23日收到)

# 肿瘤坏死因子检测技术的进展

林 缨

(中国药品生物制品检定所 生化药品与基因工程药物室 北京 100050)

肿瘤坏死因子(Tumor Necrosis Factor, TNF)是一种具有多种生物学活性的细胞因子,它参与机体免疫系统的功能调节,并在感染性疾病的发生、组织损伤和炎症等诸多方面起重要的介质作用<sup>[1]</sup>。研究资料表明血清中TNF浓度的检测有助于对疾病的分期诊断和预后<sup>[2]</sup>。另一方面,TNF对多种肿瘤细胞都具有直接的杀伤作用,很有希望成为未来治疗肿瘤的重要生物制剂之一<sup>[3]</sup>。随着研究工作的不断深入,对TNF的浓度和活性进行准确检测便愈发显得重要。因此对TNF检测技术进行全面深入的了解是十分必要和有益的,这也将有助于对TNF生理和病理作用机制的研究以及TNF制品的纯化和生产。

TNF活性的检测可分为体内测定法和体外测定法两类。

## 1 体内测定法

TNF研究的初期阶段所采用的就是体内法,它是根据TNF使小鼠移植性肿瘤产生出血性坏死的程度来判定TNF的活性<sup>[4]</sup>。这只是一种定性或半定量的方法,判断比较粗糙,很快就被体外法所取代了。

## 2 体外测定法

TNF体外检测法的实验因素要比体内法易于控制,结果比较准确、客观,重复性、相关性好。体外法根据原理不同,又可分为以下几种。

### 2.1 生物学检测方法

TNF的生物学检测方法中应用最广泛的是细胞毒性法,另外还可根据TNF的其它生物学作用,如抑制脂蛋白酶、抗病毒及刺激前列腺素的产生来检测TNF活性<sup>[5,6]</sup>。但TNF的这些作用易受其它生物学