

中药质量控制实验设计研究

赫炎¹, 沈欣¹, 靳卫东², 邵丽苹³ (1 中国中医研究院中药研究所, 北京 100700;
2 首都医科大学附属北京红十字朝阳医院, 北京 100020; 3 洛阳市第一中医院, 洛阳 471000)

摘要: 研究用多元统计分析设计, 采集和分析中药质量控制实验研究数据。结合中药研究的背景问题, 介绍了多评估指标的方差回归分析(MANOVA; multivariate regression), 及谱分析, 组成成分分析(principle analysis)和典则变量分析(canonical analysis)等数据压缩技术。提出了通过对多个药效指标的合理定量综合, 建立完整的功效指标, 并以此筛选能够综合反映药用物质基础的可测实验数据, 表达中药制品的质量。

关键词: 多元统计分析; 质量控制; 指纹图谱

中图分类号: R284, R285.5 文献标识码: A 文章编号: 1005-9903(1999)06-0026-04

An Introduction in Applications of Multivariate Analysis to Quality Control for Chinese Herbals

HE Yan¹, SHEN Xin¹, JIN Wei-dong², SHAO Li-ping³

(1 Inst of Chinese Materia Medica, China Academy of TCM, Beijing 100700;

2 Beijing Chaoyang Red-cross Hospital Affiliated to Beijing Capital Medical University, Beijing 100020; 3 The First TCM Hospital of Luo Yang, Luo Yang 471000)

Abstract: In many cases, herbal products were not well characterized and chemically identified. Therefore, how to explore the nonpurified chemical experimental data remains to be challenged. In this paper, multivariate variance analysis, multivariate regression, canonical analysis and principle analysis were introduced to experimental research for quality controls with the detailed illustrations in the background of herbal research.

Key words: multivariate analysis, quality controls, fingerprintings

中药质量研究是药理、化学和工艺学的综合研究。由于具体学科不同, 实验设计及采集实验数据的各自特点和性质不同, 从应用角度, 要求数据分析提供适当的综合分析技术。本文面对众多中药药用化学成分尚属未知, 药品质量无法用标记成分控制的实际情况, 围绕着如何运用即能反应药理作用又能综合代表药用物质基础的实验数据, 完整评估表达药物质量的实验研究设计, 提出一些综合药效、化学及工艺学的实验数据分析方法, 以期拓宽中药质量控制的定量手段。

1 指标数量化

中药实验研究中, 一些数据(成分含量、药效指标)直接采用数值表达, 且本身在研究

问题中又具有比较直观的实际意义, 可以直接被用于计算分析。但是还有一些如图谱、定性反应及采用具体加工的方法等数据必须转变成数值形式, 方能进行定量分析, 因此存在着如何在尽量不损失信息的情况下, 对其数量化的问题。现介绍两种数量化的方法。

1.1 定性数据的数量化 设某一定性指标用阴性或阳性表达, 则数量化后取 $y = \begin{cases} 1, & \text{阳性} \\ 0, & \text{否} \end{cases}$, 例如在毒性实验中, 以数值变量 y 表达小鼠死亡与否, 则 $y = \begin{cases} 1, & \text{死亡} \\ 0, & \text{否} \end{cases}$ 这种数量化的特点是当 y 为 0, 1 两态变量时, y 是对某一随机事件发生概率的直接记录。例如在毒性实验中, 当 y 取 1 时, 表示某一小鼠

死亡随机事件的发生概率观测值为1。并且这种0,1的取值,无论从相对意义和绝对意义都是对称的。同时 y 的均值本身就是对某一随机事件发生概率的估计值。

1.2 图谱数据的数量化 在有效成分不清楚的中药质量控制研究中,由于研究目的是寻找能完整反映药效、应用方便的客观质控指标,因此,如果能将标记成分的概念扩充到标记指标(如采用理化性质等指标)而避开暂时未知的活性成分,应是质控研究的重要部分。而这些理化指标往往都是以图谱形式表达。因此存在一个谱分析的问题。即通过对

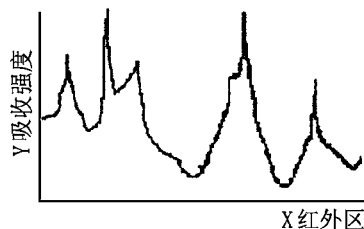


图1

图谱的数值规律的定量分析以表达其数值特征(代表理化性质和药物含量的标记指标)。下面介绍一种常用谱分析

方法。某一药物在红外区的吸收构成连续图谱图1。根据谱叠加原理, $y = \sum \cos(\alpha_j + \beta_j x)$; 或 $y = \sum \sin(\alpha_j + \beta_j x)$; $j = 1, 2, \dots, m$; 其中的 α_j 和 β_j 为图谱的数值特征, m 根据具体情况而取定。如此,我们即获得了可以表达图谱特征的一组数值群。从而通过分析这些数值特征同药品质量指标(如药效)之间的定量关系,有望建立可控质量标准。

2 数据的综合分析

综合分析,是指在某一综合定量原则下,运用特定的算法对某一样品的属性所作评价的过程。例如,通过药品的一些理化性质对其相应的制品质量给出评价。又如,通过一系列的症候群改变情况对疗效给出评估。

这些分析主要涉及到方差和回归分析等。其中,药效指标仅为一种时(univariate),相应的分析即为常用的单变量方差分析(ANOVA)和多从回归(multiple regression)。同时,根据实验控制因素的数量,又分“单因素和多因素”方差分析。当药效指标多于一种时(multivariate),相应的分析为多变量方差分析

(MANOVA)和多元回归(multivariate regression)。当需要用多种药效指标同时综合定量时,又可用典则变量分析(canonical analysis)。下面介绍几种相关的多变量统计分析方法。

2.1 单变量方差分析 具体做法是通过在考察因素的不同水平(如不同温度),一种观测指标值的变化与误差变异项的比较,决定其与考察因素的相关性。由于应用者往往愿意采用一些系统抽样的实验设计技术,如拉丁方以及正交表,在保证各因素之间能够相互正交的同时,对每个因素的独立作用作出评价。因而,在实验因素的筛选中得到了广泛的应用。有关这类研究文献报告较多,笔者就不罗嗦。在此只是为了能够使应用者在利用正交设计所进行优化研究时,对分析结果所具有的优良性,不加证明地直接给出。在简单随机样本中, y_1, \dots, y_n 中,最大值 y_n 值大于总体中占 $n/(n+1)$ 比例的其他 y 值。而对同样大小的系统样本,在一般情况下,其最大值至少不比简单随机样本中的 y_n 小。例如应用这一估计,在对4因子3水平的试验,利用正交表 $L_9(3)^4$ 实施9次试验,优化结果意义如下:9次试验中,所得的最大值在 $3^4 = 81$ 个全部水平组合中,将优于占比例9/10的其他测试值;即在所有可能的81个组合条件中,一般要名列前9名。

2.2 多变量方差分析 作为方差分析,多变量同单变量分析的基本思想是相同的。但是由于这种分析,在一次性分析中可考虑各因素不同条件下的多种观测指标变化,因此它可通过分离观测指标之间的协方差而高效率地检测出考查因素是否同时影响这些观测指标,因而筛选那些可全方位影响药品质量的可控因素。同时这种分析还可以避免分别逐次用单一指标评估考查因素的方差分析造成的由于重复利用小概率事件获得的假阳性结论^[1]。因此,尤为适合中药质控实验研究。

2.3 回归分析 药品质量是由多种实验可

控制因素决定, 因此质量可作为因变量 Y , 可控制因素作为自变量 X , 假设希望在可能的情况下, 用

$$Y_i = \sum \beta_j X_j + \epsilon, i = 1, \dots, n; j = 1, \dots, m \quad (1)$$

的线性模型拟合(simulation)的数据同实际实验数据比较, 来检验该模型的可取性以及在此模型中各自变量对因变量 Y 的作用。其中 Y 是单变量时, 为多重回归, Y 是多变量时为多变量回归。并期望以 Y 筛选与药品质量关联较大的因素或工艺参数。这是一个典型回归问题。实际应用时, 往往由于某个因素不仅同药物质量 Y 有关联, 同时同其他因子 x 之间有相关(multicollinearity), 因此, 对线性模型优化时(如工艺参数的确定), 为了提高模型的价值(增大 F 值), 从应用角度, 一些因子需被剔除。近年来, 基于这种意义的线性有偏估计方法有了很大发展, 使得对实验取样设计的要求有所宽松。在质量控制研究中得到了广泛应用。

2.4 典则变量分析 中药质量控制研究中, 饮片功效往往需要从多方面进行评估。例如对清热化痰药, 其功效至少应从清热和化痰两个指标给予评估。运用上述的多变量方差及回归分析仅仅可以高效率地检查出哪些实验可测数据(考察因子)同这两个指标相关及每个指标同这些因子的定量关系, 从而为以这些因子作为可控的质量标准提供定量依据。但如何综合评估指标 $Y(Y' = (y_1, y_2))$, 进而筛选能够综合表达制品功效 Y 的可控因子 X , 仍未解决。例如指标 y_1 反映清热, y_2 反映化痰, 那么如何综合这两个指标呢? 简单地将 y_1 与 y_2 相加, 然后利用多重回归建立 $Y_1 + Y_2$ 与 X 的定量关系, 这种一视同仁的做法, 显然有悖于通过清热而化痰的理论。其道理正如混合含有不同数量土豆与苹果的两堆食品, 其价值不能简单用总的数量评估一样。典则变量分析即是分析一组因变量 Y 同时与一组自变量 X 之间的相关程度以及建立相应的定量关系: $\alpha_1 y_1 + \alpha_2 y_2 = \beta_1 x_1 + \beta_2 x_2 +$

$\dots + \beta_n x_n$ (1 式), 从而给 y_1 与 y_2 合理的一种加权方法。其中 α_1, α_2 分别为 y_1 与 y_2 的加权, 其参数 $\alpha_1, \alpha_2, \beta_1, \beta_2, \dots, \beta_n$ 的寻找确定原则是根据使得 1 式中 Y 与 X 相关最大。因此, 它非常适合中药质量评估分析。

2.5 主成分分析(principle analysis) 我们通过举例说明这种分析的基本思想。例某中药饮片可用于治疗两种不同临床病症。由于药材活性成分还不清楚, 实验研究试图从药物水提物的紫外吸收图谱(假设在紫外有吸收), 饮片粉末色度和水提取收率等数据中, 寻找质控指标, 并以此建立两种互不相关的综合质量标准, 作为该种饮片对两种临床应用的质控标准。按多元分析设计应测得如下实验数据(表 1)。采用主成分分析, 对 X 数据阵线性变换, 找出 6 个相互独立的主成分(principle component)。由于这 6 个正交关系的主成分的离散度可由 X 数据阵的协方差矩阵或相关矩阵的特征根代表。因此, 根据其特征根大小, 我们可以直接选出离散度最大的前两个主成分 Z_1 和 Z_2 (直观理解, 主成分 Z_2 离散度越大, 对同一样品在 X 空间, 用 Z_2 测度其分辨率越高)。 $Z_1 = \sum \beta_i X_i; Z_2 =$

表 1 X 数据库

样品号	色度值		紫外吸收图谱		水提取收率		
	X_1	X_2	X_3	X_4	X_5	X_6	
1	~	~	~	~	~	~	
2	~	~	~	~	~	~	
3	~	~	~	~	~	~	
4	~	~	~	~	~	~	
5	~	~	~	~	~	~	
6	~	~	~	~	~	~	= X
7	-	-	-	-	-	-	
8	~	~	~	~	~	~	
9	~	~	~	~	~	~	
10	~	~	~	~	~	~	
11	~	~	~	~	~	~	

$\sum \beta_i X_i$ 。 Z_1 和 Z_2 的实际意义为同一样品根据不同应用目的, 所建立的两种运用相同因子但不同打分标准的综合得分。由于两者不相关, 如同一制品用 Z_1 评价可能完全不同于用 Z_2 所得结果。就是说同一制品对不同病症具有不同质量标准。

在具体计算过程中, 我们还可以通过 X 数据阵计算得出 Z_i 与 X_i 的复相关系数 (multiple correlation), 从而剔除一些对主成分贡献较小的变量 X , 达到压缩数据的目的。

例如, 经过数据压缩后, 得 $Z_1 = \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$; $Z_2 = \alpha_3 X_3 + \alpha_6 X_6$; 说明主成分 Z_1 与紫外吸收图谱的两个参数 X_4 、 X_5 及水提物收率 X_6 有关; 紫外吸收图谱的另一个参数 X_3 和水提物收率 X_6 与 Z_2 有关。应当指出的是上面的方法无论用在分类还是聚类问题, 只是对样本在研究范围 (X 空间) 的数值规律总结, 都有待于实践的验证。因此往往需要第二阶段验证研究。

2.6 判别分析 (discriminate analysis) 上面的例子是在对饮片质量缺乏具有权威性评价的情况下, 如何利用可观测到的指标在样品身上反应的数值规律建立评价标准。在实际应用中, 我们还常常遇到相反的情况, 即药品的分类或质量评价可由权威人士或公认的实验研究结果评价确定。例如某药材的分类可由专家认定、饮片质量可通过动物或临床实验给予客观地评价。但出于应用的考虑, 我们更感兴趣是能否根据这些评价的标准, 找出一些与其有相关性的简便可行指标来替代上面的评价方法。判别分析就是根据这些有评价基准数据的样本, 分析找出与评价数据

有相关性的因子, 最终建立判别标准 (判别函数), 并以此对未知分类的样品作出评价的多因素统计分析方法。应用判别分析方法, 我们可以通过以某一公认指标 (如活性成分含量) 对一些简便可测的理化指标进行多重回归, 并应用方差回归显著性检验的技术对这些指标进行筛选。最终给出判别函数 $y = \sum \beta_i X_i$: 通过对训练样本回代建立适当 y 界值, 对未知样品的质量作出评价。当然判别分析的判别标准很多, 具体应用时, 我们可以根据计算机软件包的帮助说明, 选定算法, 进行实验设计和分析。

3 小结

本文对中药质量控制实验研究数据的多变量的多因素综合分析设计及方法的具体思想作了简单介绍。特别是针对中药很可能是通过多成分综合作用, 而我们对活性成分又不清楚的情况下, 如何利用多元统计分析解决中药质量控制问题; 提出了标记指标的概念及具体应该采用的算法。应当指出的是西方对植物药品的新药申报审批, 指纹图谱等能综合反映药用物质基础的实验数据以及相应的多元分析, 已被采用和日益重视^[2]。本文作为抛砖引玉, 希望早日在国内出现这方面应用实例, 并得到同行的认同。

参考文献:

- [1] 赫炎, 沈欣. 论中药实验研究数据分析[J]. 中国实验方剂学杂志, 1999, 5(2): 57
- [2] FDA[M]. Guidance on Botanical Drug Products, 1996. 16

(收稿日期: 1999-03-16)