

•思路与方法•

数理统计学在中药提取规范化中的运用

王建明^{1*}, 王和平², 于鹤丹¹

(1. 黑龙江中医药大学, 黑龙江 哈尔滨 150040; 2. 浙江医药高等专科学校, 浙江 宁波 315100)

[摘要] 目的: 运用数理统计学方法使中药提取工艺规范化和数据化, 为中药提取过程标准化(GEP)的建设提供参考。方法: 采用理论分析与实例介绍相结合的方法, 分别从正交试验设计法的应用, 均匀设计实验法的应用及多指标综合加权评分法的运用三个方面阐述了数理统计学在中药提取规范化即 GEP 实施中的运用。结果: 通过运用数理统计学可以获得最佳的和最合理的提取工艺条件使之达到规范化、科学化、数据化。结论: 在实施 GEP 中运用数理统计学方法是必要的和重要的。

[关键词] 正交试验设计; 均匀设计试验; 综合评分; 中药提取规范

[中图分类号] R283.6 [文献标识码] A [文章编号] 1005-9903(2007)05-0059-04

Utilizing of Mathematical Statistics in the Standardization of Extracting Chinese Medicine

WANG Jian-ming^{1*}, WANG He-ping², YU He-dan¹

(1. Heilongjiang University of Chinese Medicine, Harbin 150040, China;

2. Zhejiang Pharmaceutical College, Ningbo 315100, China)

[Abstract] **Objective:** To use the method of mathematical statistics for standardization and digitalization in extraction of herbal medicine that providing reference for the construction of GEP. **Method:** The method of theory analyses with examples was used respectively on the utilization of orthogonal experiment, homogeneous design and synthesizing evaluation of multiple indexes in extracting standardization of chinese medicine as GEP. **Result:** The conditions of extracting Chinese medicine were obtained by utilizing mathematical statistics. **Conclusion:** To utilizing mathematical statistics in realizing GEP is essential and important.

[Key words] orthogonal experiment; homogeneous design; synthesizing evaluation; extracting standardization of chinese medicine

无论是提取工艺路线还是精制工艺路线, 或是其它工艺路线, 都需要应用最佳工艺条件, 实施规范化操作。经过对处方的科学分析, 做出正确的、科学的药材处理路线和提取、纯化工艺路线的设计之后, 必须进行最佳条件的优选。最佳条件的优选离不开数理统计学。

随着科学技术的进步和各学科理论知识的相互

渗透, 利用现代数学统计学方法优化试验设计已经成为中药研究的重要领域。按照数学统计学方法设计一系列试验, 从而得出某种科学的指标, 优化出中药提取最佳条件。当前中药提取的工艺要求规范化, 即以 GEP 为目标^[1,2], 因此所有工艺设计都要经过科学的实验筛选。这就是利用数学方法, 结合药剂学理论和提取、纯化方法, 事先确定考察因素及因素的几种状态, 设计出试验次数少、成本低且又具有一般性的试验方案, 从而筛选出最佳工艺条件。目前我们通常应用的是正交试验设计和均匀设计实验。

[收稿日期] 2006-07-17

[通讯作者] * 王建明, Tel: (0451) 82196262; E-mail:

Wangjianming@hljucm.net

1 正交试验设计法的应用

在中药提取和药物生产过程中,总希望要做到数量多、时间快、质量好、消耗少地完成试验。这就需要通过多次实验来优选最佳条件,在优选最佳条件的过程中,如果试验设计做得好,即使实验次数不多,也可以得到满意的结果;反之,如果试验设计得不好,试验次数多也不能获得满意结果。尤其在考察的条件(即因素,如温度、时间、浓度、pH 值、次数、粒径、溶剂量、方法等)较多,并且每个条件又有若干等级(即所谓的水平)的情况下,怎样合理地设计试验是很值得研究的一个问题。正交试验设计就是一个处理多因素、多水平试验的一种科学方法^[3,4]。

正交试验设计的关键是表头设计,事先必须理清所考察的因素数和各因素的水平数,水平相等的表头设计可以查一般的正交试验设计表,水平不等的表头要查混水平正交设计表。对于各因素间有无相互作用也在考虑之内,若有相互作用的须用交互作用正交表。正交试验设计可以有多个考察指标,一般单指标的数据处理方法用方差分析法,多指标的数据处理方法可用综合平衡或综合评分法处理。

我们在中药提取工艺学研究中常用的正交表有两种。

其一是正交表 $L_8(2^7)$,即水平数为 2,因素最多可安排 7 个,试验次数是 8 次;其二是正交表 $L_9(3^4)$,即水平数为 3,因素最多可以安排 4 个,试验次数是 9 次。各表的因素数可以选满,也可以有空白。试验的考察指标可以是化学指标(如有效成分含量),可以是物理指标(如固形物总量),或是药理效应指标(如作用强度);可以是单指标,也可以是二个以上指标。

对于考察结果的判断有二种方法:①直观分析。可以从极差的大小来确定因素的主次,用各水平数据之和的大小来确定水平的优劣。极差为某因素下水平数据之和的最大值减去最小值而得,极差越大说明该因素影响越大,即为主要因素。水平数据之和较大者为优选条件。②方差分析。为避免实验误差因素,可选用 F 检验的方差分析确证因素的水平变化所引起的实验结果的变差大大超过实验误差。但若方差分析结果没有显著性差异时,则可从极差分析中选择最佳工艺条件,只能说明因素的水平变化对工艺影响较小。

例如用正交试验法研究三七提取工艺中^[5],试

验的考察指标为测定每个样品中的三七总皂苷量。

在对溶媒和提取方法初步选择以后,选定乙醇浓度、提取方法和乙醇用量作为考察的 3 个因素,每个因素各取 3 个水平。试验采用 $L_9(3^4)$ 表头进行正交试验设计,试验结果通过三七总皂苷含量加以表达,并计算极差。

通过极差分析可以看出,影响最大的因素是提取方法,以渗漉法为最好;其次是乙醇浓度,以 75% 为最好;乙醇用量影响最小,以 10 倍量为最佳。直观分析结果以 $A_2B_1C_2$ 为最佳工艺。

为进一步确证,将数据进行方差分析。结果与直观分析相一致。最佳工艺为采用渗漉法、75% 乙醇收集 10 倍量体积。通过科学的数据处理,使三七提取工艺规范化,数据化。

2 均匀设计实验法的应用^[3,6]

均匀设计是一种多因素试验设计方法,所谓试验设计是指对于即将进行的试验进行合理的安排,使之用尽量少的试验达到预期的目的。

均匀设计的基本思想就是抛开正交设计中“整齐可比”性的特点而只考虑试验点的“均匀分散”性,即让试验点在所考察的试验范围内尽量均匀地分布,“均匀设计”由此而来。由于不再考虑“整齐可比”性,那些在正交设计中为整齐可比而设置的实验点可以不再考虑,因而用均匀设计安排试验可以大大减少试验次数,且试验次数与各因素所取的水平数相等。用均匀设计可适当增加试验的水平数,而不必担心导致象正交设计那样其试验次数呈平方次增长的现象。

为了达到均匀布点的目的,和正交设计类似,可以使用均匀设计表(简称均匀表)安排试验。该设计对水平数较大的试验,优势更为突出,其试验次数仅需与水平数相当,最多比水平数多一次。均匀表的表头形式是 $U_n(t^s)$, U 表示均匀表, n 表示试验次数, t 表示水平数, s 表示最多可安排的因素数。这里 $n = t, s = t - 1$,即试验次数与所取水平数相等,最多可安排的因素比水平数少 1。例如 $U_5(5^4)$ 表示一个均匀表,最多可安排 4 个因素,每个因素分 5 个水平,做 5 次试验即可。

常用的均匀设计表有 $U_5(5^4)$ 、 $U_7(7^6)$ 、 $U_9(9^8)$ 、 $U_{11}(11^{10})$ 及 $U_{13}(13^{12})$,最多可以安排 13 个水平表设计实验。当因素的数目未达到所用均匀表的限数时,可根据因素数查相应的使用表来确定选用的列

号安排试验。

关于均匀设计的数据处理可按以下要点加以掌握:

(1) 将试验结果与各水平归一化处理, 即各因素归一化处理 = (实际值 - 最小值) / (最大值 - 最小值)

(2) 采用最小二乘法进行方程拟合, 可获多元回归方程, 即

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

其中 $b_0, b_1, b_2, b_3, \dots, b_n$ 为多元回归方程之系数, 对应于变量 X (因素) 的序号。

如果某个系数与其它系数相比非常小时, 说明该因素(变量)对试验影响不大, 可以舍去而重新对其它变量进行回归。

(3) 结果判断

可有以下三点: ①系数 b 为正值时, 表明该因素对试验为正相关, 数值越大越好, 数值越大表明该因素影响越大; ②系数 b 为零时, 影响可以忽略, 可根据实际生产情况选择最方便的水平; ③系数 b 为负值时, 表明该因素对试验为负相关, 数值越小越好, 可选该因素的最小水平。

(4) 优化条件的确定

最佳条件的确定可以用求函数最大值的方法计算出各因素的最佳值。但是理论计算的最大收率(或最好结果)实际实验有时是达不到的。

例如用均匀设计法优选侧柏叶中槲皮苷的提取纯化工艺中^[7], 通过预试验确定了以渗漉法为试验的提取方法, 试验考察指标为槲皮苷的含量。

首先用乙醇浓度、乙醇用量、浸泡时间 3 个因素考察提取工艺, 每个因素设 5 个水平, 通过试验获最佳提取条件。

然后根据所得最优条件, 选择加水量、加 NaCl 量、调配 pH 值及萃取溶剂体积 4 个因素, 每个因素 5 个水平, 再次用均匀设计考察纯化工艺的最优条件。提取与纯化的因素与水平, 按 $U_5(5^4)$ 使用表安排试验, 并获得测定数据。

将提取数据通过多元非线性回归, 获回归方程:

$$Y = -0.41915 + 0.013672x_1 + 0.002502x_2 + 0.001093x_3 - 6.9 \times 10^{-5}x_1^2 - 4.8 \times 10^{-5}x_2^2 + 1.49 \times 10^{-5}x_3^2 - 4.6 \times 10^{-5}x_1x_3 \quad (1)$$

用求多元函数最大值的方法得到了槲皮苷含量最高的实验条件为乙醇浓度 73%, 浸泡时间 26h, 乙

醇用量 77 mL。含量最优值为 $y = 0.158\%$ 。

对最优纯化条件结果进行多元线性回归获回归方程(2):

$$Y = 0.3588 - 0.00257x_1 + 0.000432x_2 - 0.00533x_3 - 0.00105x_4 \quad (2)$$

通过方差分析可知方程无显著意义, 再进行逐步回归(去掉氯化钠因素)又获回归方程(3):

$$Y = 0.348981 - 0.00238x_1 - 0.00495x_3 - 0.00097x_4 \quad (3)$$

通过方差分析该方程仍无显著意义, 通过计算准回归系数知 pH 值对含量影响不大, 第三次回归得方程(4):

$$Y = 0.323541 - 0.002236x_1 - 0.000896x_4 \quad (4)$$

方差分析方程有显著意义。

结果证明, 在所考察的实验条件内减少用水量 and 减少萃取溶剂用量能使槲皮苷含量增加。若提取物加水 20 ml 溶解, 用乙酸乙酯萃取 6 次, 每次 30 mL, 可使槲皮苷含量达到 0.2519%。

3 多指标综合加权评分法的运用

在很多情况下, 为取得更为科学, 更为规范的提取工艺, 往往采用二个以上的多指标加以考察, 因此为使各因素水平的贡献具有均等的机会, 必须对各指标数据加以加权处理, 进行综合评分^[8]。

例如红豆杉细胞培养中以木质素与紫杉醇二个指标综合评价时^[9], 木质素含量为 10 位数量级(如 5 号样品为 15.550), 而紫杉醇含量极低是小数点后 1 位(同为 5 号样品是 0.254); 另一例是麦冬和枇杷叶的混合提取中以水提醇沉物重及熊果酸含量为指标综合评价时^[10] 同样前者为个位数量级(如 5 号样品为 4.84), 而熊果酸含量也是小数点后一位(同为 5 号样品是 0.487)。此二例中如果把二指标直接相加的话, 则前者会掩盖后者的贡献, 也就不能正确反映优化结果。因此要想使二者数据对本次提取的贡献均等, 则应把后者加权处理使其上升为同等地位, 第一例中紫杉醇含量则予以乘以 100 处理使其达到 10 位数量级, 第二例中熊果酸含量则予以乘以 10 处理使其达到个位数量级, 然后再将二个指标数值相加方能得到均等贡献的评价, 才能进一步进行有效的统计计算, 从而得到科学而规范的优化提取条件。

应该指出: ①综合加权评分的方法非常灵活, 要根据指标性质研究评分方案。有时也可采取倒数的方法评出单相关。②本例属于定量多指标试验, 如

果确定的是定性指标,如颜色、气味等,可把各号样品试验结果按规定指标及其重要程度相互比较,把最好的试验结果定为100分,其它样品酌情扣分处理。③如果试验结果为越小越好,可用减法转化分数。若为结果越大越好(如有效成分含量)则可用加法,例如所有样品中测得最高含量为45 μg ,则用 $Y_1 = 55 + Y_1$ 的公式计算,即 $(55 + 45) = 100$ 分。以此类推,其余各样品综合评分均小于100分。

以上从三个方面介绍了在中药提取工艺规范化,即GEP实施中数理统计学的运用,都是一些普及的内容,但却显示了较之比较试验法更优越、更科学之处。只有充分而灵活地运用数理统计学于中药提取工艺设计之中,才能实现GEP所期待的目标,才能奠定好中药现代化实现的基础。

[参考文献]

- [1] 史道华. 中药现代化的几个关键问题及其对策[J]. 中草药, 2004, 35(9): 961.
- [2] 孙成逊. 中药提取过程标准化(GEP)与药品质量的关系[J]. 中成药, 2006, 23(2): 276.

- [3] 平其能. 现代药剂学[M]. 北京: 中国医药科技出版社, 2001, 215-231.
- [4] 刘婉如, 庞善起, 张千里, 等. 关于正交设计与均匀设计的比较[J]. 数理统计与管理, 1995, 14(2): 41-46.
- [5] 唐红芳, 毛丽珍, 徐世芳. 正交试验法研究三七提取工艺[J]. 中草药, 2001, 32(1): 26-28.
- [6] 王鹏, 王玉珠, 沈建民. 均匀设计及其在药学中的应用[J]. 沈阳药学院学报, 1989, 6(4): 297-306.
- [7] 曾詮, 易登录, 吴洪元, 等. 用均匀设计法优选侧柏叶中槲皮苷的提取纯化工艺[J]. 中草药, 1998, 29(12): 805-807.
- [8] 孙振球. 医药统计学[M]. 北京: 人民卫生出版社, 2002. 181-186.
- [9] 赵应征, 鲁肇涛, 梅兴国. 常用多指标综合评价法在优选实验中的应用[J]. 医学研究生学报, 2004, 17(7): 624-626.
- [10] 李矗, 唐丽琴, 刘圣, 等. 混合指标的正交设计法优选麦冬、枇杷叶的复方提取工艺的实验研究[J]. 安徽医药, 2003, 7(4): 245-248.