

文献中正交试验的常见问题分析和解决方法

张玉军¹, 莫志江^{2*} (1.广西壮族自治区药品和医疗器械集中采购服务中心, 南宁 530021; 2.广西壮族自治区人民医院药学部, 南宁 530021)

摘要: 目的 提高正交试验的质量和实用性。方法 对文献报道的正交试验进行总结分析。结果 部分正交试验存在的问题和后果是: 无空白列导致夸大随机效应; 无理由忽略交互作用导致混杂现象; 手工计算导致方差分析错误; 最优工艺不考虑成本因素导致无谓增加生产成本。结论 正交试验应该留有空白列; 否定交互作用的存在应该有实验或文献的支持; 应该使用专业统计软件分析数据以确保结果的准确性; 工艺考察应包括成本-效益评估。

关键词: 正交试验; 方差分析; 混杂现象; 成本-效益

中图分类号: R91

文献标志码: A

文章编号: 1007-7693(2013)06-0696-05

Common Mistakes Made in Some Literatures When Performing Orthogonal Experiments and Solutions

ZHANG Yujun¹, MO Zhijiang^{2*} (1.Medicines and Medical Equipment Centralized Purchasing Service Center of Guangxi Zhuang Autonomous Region, Nanning 530021, China; 2.Pharmacy Department, the People's Hospital of Guangxi Zhuang Autonomous Region, Nanning 530021, China)

ABSTRACT: OBJECTIVE To improve the quality and utility of orthogonal experiments. **METHODS** The orthogonal experiments in published literatures were reviewed. **RESULTS** The common mistakes made in some literatures and outcomes were: orthogonal design or orthogonal table without blank column, resulting in inflating the random effect; baselessly ignoring the interactions, resulting in confounding bias; wrong results of analysis of variance (ANOVA) because of manual calculation; the selection of optimal process without considering cost-effectiveness, incurring cost increase of manufacturing. **CONCLUSION** Blank column of orthogonal table should be reserved for error estimated; the interactions could be ignored only when their existence are denied in relevant experiments and literatures; the data of orthogonal experiments should be analyzed using statistical software to make sure the accuracy of the results, and the cost-effectiveness should be considered in the selection of optimal process.

KEY WORDS: orthogonal experiment; analysis of variance(ANOVA); confounding bias; cost-effectiveness

与析因试验(factorial experiment)需要考虑各个因素各个水平的全部组合(或称搭配)不同, 正交试验(orthogonal experiment)仅仅考虑一部分主要的组合, 因此大幅减少试验次数, 是药学工作中常用的实验方案。但是, 国内药学文献中的一些正交试验在实验设计和数据分析中存在较多不足和错误, 现就一些常见问题进行详细分析, 期望能提高正交试验的质量和实用性。

1 材料与方法

1.1 数据来源

为直观地显示各种常见错误的原因, 笔者主要以一个寻找复方大黄制剂最佳工艺的模拟实验来进行说明。假设 4 个试验因素分别为 A、B、C、D,

每个因素均为 3 水平, 同时假设需要考察的指标为大黄素含量和浸膏收率(为便于分析, 附上模拟实验结果)。对这类实验, 许多文献报道采用 L₉(3⁴)的正交设计, 见表 1, 这类设计不但常见于普通药学期刊^[1-2], 核心类药学期刊也是屡见不鲜^[3-8], 也由此带来一系列的典型问题。

1.2 数据处理

国内文献绝大多数使用手工计算(包括借助计算器或电脑, 但未利用专业统计软件的方法)和直观分析, 这里不再赘述。为保证结果的准确性, 本研究使用 SPSS 15.0 for Windows 软件中的一般线性模型(general linear model, GLM)处理全部数据。

基金项目: 广西壮族自治区卫生厅课题(Z2010256)

作者简介: 张玉军, 男, 硕士生, 副主任药师 Tel: (0771)2801500
Tel: (0771)2186108 E-mail: mozhijiang89@163.com

E-mail: zhangyj20@126.com *通信作者: 莫志江, 男, 主任药师

表 1 一些文献 L9(3⁴)正交表的常见设计和模拟实验结果
Tab 1 Common layout of L9(3⁴) on some literatures and hypothetical results

列号	1	2	3	4	大黄素 含量/ mg·g ⁻¹	浸膏 收率/ %
因素	A	B	C	D		
试验号	交互作用 ¹⁾	(C×D) ₁	(C×D) ₂	(A×B) ₁	(A×B) ₂	
1		1	1	1	8.54	21.1
2		1	2	2	8.28	20.0
3		1	3	3	8.24	19.7
4		2	1	2	8.32	20.2
5		2	2	3	8.15	23.7
6		2	3	1	7.88	22.0
7		3	1	3	7.60	24.1
8		3	2	1	8.42	21.2
9		3	3	2	8.01	23.5

注：¹⁾交互作用的解释见下文
 Note: ¹⁾Interaction, discussed below

2 问题分析和解决方案

2.1 未预留空白列

误差项(error)被用于探测随机效应所造成的结果变化，而判断各种试验因素是否对实验结果产生明显影响需要与随机效应(误差项)进行比较，因此应该预留空白列用于计算单纯的随机效应。但是，表 1 的 L9(3⁴)正交表被排满 4 个试验因素而未预留空白列，这在很多情况下将导致随后的方差分析结果不可靠。使用 GLM 处理表 1 的数据，方差分析结果见表 2 和表 3。

表 2 大黄素含量的方差分析结果

Tab 2 Results of ANOVA for emodin

因素	离均差平方和(SS)	自由度	均方(MS)	F 值	P 值
A	0.185 3	2	0.092 6	2.15	0.319
B(误差)	0.086 6	2	0.043 3		
C	0.128 9	2	0.064 4	1.45	0.402
D	0.272 3	2	0.136 1	3.14	0.241

表 3 浸膏收率的方差分析结果

Tab 3 Results of ANOVA for extract rate

因素	离均差平方和(SS)	自由度	均方(MS)	F 值	P 值
A	10.936	2	5.468	259.00	0.004
B(误差)	0.042 222	2	0.021 111		
C	2.782	2	1.391	65.89	0.015
D	9.076	2	4.538	214.95	0.005

在方差分析中，F 值的计算公式为

$$F = \frac{MS_{\text{因素}}}{MS_{\text{误差}}} \tag{1}$$

$$\text{而 } MS = \frac{SS}{\text{自由度}} \tag{2}$$

如果预留任意一列作为误差项，则该列的结果变化(离均差平方和，SS)将完全来自于随机效应。但是，由于表 1 在之前的正交设计中未预留空白列作为误差项，在随后的方差分析中只能把对结果影响最小(SS 最小)的 B 因素指定为误差项，此时误差项的结果变化来自于 B 因素和随机效应 2 个方面，而不是单纯的随机效应。此时，若 B 因素对结果的影响明显高于单纯的随机效应，则 MS_{误差}将会偏大[按(2)计算]，造成 F 值偏低[按(1)计算]；若 B 因素对实验结果的影响与随机效应基本相当则不受影响。

对于表 2 的结果(全部因素无显著性)，4 个因素 A、B、C、D 的 SS 相差不大，此时将 B 因素指定为误差项，由于不知道真正的随机效应有多大，A、C、D 所对应的低 F 值完全有可能是因为 B 因素的影响明显大于随机效应所致，这样的结果显然不可靠。

对于表 3 的结果(所有因素均有显著性)，由于 B 因素的 SS 明显小于其他任何一个因素，此时 A、C、D 所对应的 F 值已经很大，因此最主要的因素已经被确定，而 B 因素的影响力相对来说太小，不管相对于单纯的随机效应是否具有显著性，将其忽略也不明显影响结果，所得结果仍然比较可靠。值得指出的是，表 3 的结果是这种未预留空白列的设计方案所能成立的唯一情形，其他情形(全部或部分因素无显著性)都不完全可靠，其中以表 2(全部因素无显著性)的结果最不可靠。

因此，不应采用无空白列的正交设计方案，除非采用重复测量设计，此时误差项的 SS 可通过重复试验时的结果变化来计算，不必通过空白列。

2.2 忽略交互作用从而导致混杂现象

一个好的正交设计，除了需要预留空白列，还需要注意交互作用，避免出现混杂现象。

如果 A 因素所导致的结果变化与 B 因素有关，且变化明显大于随机效应，就称 A、B 两个因素之间存在交互作用，并以 A×B 表示。例如，在化学合成工艺中，需要考察催化剂(甲催化剂、乙催化剂)和反应温度(70，80，90 ℃)对产品收率的影响。假设甲、乙 2 种催化剂的最适宜温度分别是 70 ℃和 90 ℃。很明显，此时不能孤立地说哪种催化剂好，还需要考虑催化剂所搭配的反应温度。甲催

化剂配 70 ℃最好，而乙催化剂配 90 ℃最好；同理，也不能孤立地说反应温度高好还是反应温度低好，还需要看搭配何种催化剂，这种情况就被称为催化剂与反应温度之间存在交互作用。交互作用有正负之分，如药物间的协同作用属正交互作用，而拮抗作用属负交互作用。

一般情况下不应简单否定交互作用的存在，需要通过实验或查阅相关文献资料后才能确定。但是，许多文献报道没有提及忽略交互作用的理由，直接就将主因素和交互作用项安排在正交表的同一列上，这其中也不乏有些研究者不了解正交表的构造所致，这将导致所谓的混杂现象，也就是无法分清是哪种情况造成实验结果的变化，类似于表 1 未预留空白列(试验因素和误差项被安排在同一列上)。

笔者建议研究者按以下几个方法判断试验中可能存在的交互作用。

通过经验或专业理论来推测：可以结合测量指标和试验因素做一些符合专业知识的合理推测。假设表 1 的 A 为提取次数，B 为料液比，C 为醇浓度，D 为提取时间。如果测量指标是浸膏收率的话，那么 A、B、C 和 D 之间有可能不存在交互作用；但是，如果测量指标是大黄素含量的话，那么不同醇浓度有可能造成沉淀物中可吸附大黄素的成分及其比例的差异，如果这种吸附作用足够大，且与接触时间有关，那么就不应该排除 C 和 D 之间的交互作用。显然，这种推断具有一定的不确定性，证据力不够强。

通过相关文献资料来佐证：如果有相关文献报道不存在交互作用，或是依据相关文献的结果，经研究者进一步分析能够排除交互作用(方法见下文)，则可以认为不存在交互作用，证据力比较强。

通过前期实验来证实：例如，如果重点怀疑 C 和 D 之间的交互作用，可以考虑事先对 C 和 D 进行析因试验以确定 C 和 D 之间的交互作用，其结论有最强的证据力。

事后分析：可以对已经完成的正交试验(自己的试验，或文献上别人的试验)进行事后分析以判断交互作用的强弱。假设出现了表 2 的结果(假设只有 3 个因素，且全部因素均无显著性)，则不能排除交互作用的存在，因为空白列共存于交互作用列上；假设出现了表 3 的结果(假设只有 3 个因素，且全部因素均有显著性)，则可以忽视交互作

用的存在，因为尽管空白列共存于交互作用列上，但其影响力相对于主效应来说是比较弱的；余此类推，其他结果参考交互作用列出现的位置(见下文)不难推断其强弱。

每一项交互作用在正交表中至少占 1 列(与相关主因素的水平数有关)，其出现的位置与相关主因素的位置有关，已有统计学书籍把这些情况编列成表格^[9]，参考这些表格可以正确安排各个试验因素。例如，对于一个 4 因素(分别以 A、B、C 和 D 表示)2 水平的试验，如果前期实验或资料已证实 B×C 不明显，而其他交互作用不确定，则可以使用 L8(2⁷)正交表，正确的试验因素安排见表 4。

表 4 4 因素 2 水平正交试验在 L8(2⁷)表上的正确顺序
Tab 4 Proper layout of L8(2⁷) of 4 factors of 2 levels each

列 号	1	2	3	4	5	6	7
因素或交互作用	A	B	(A×B)	C	(A×C)	(B×C)	D
			(C×D)				

4 个主因素 A、B、C、D 被分别安排在第 1、2、4、7 列上，避免了可能的混杂现象，同时也可以考察部分交互作用是否显著。第 6 列为 B×C 出现的位置，由于已经证实无显著性，在随后的方差分析中可以被指定为误差项。如果将 A、B、C、D 按顺序分别安排在第 1、2、3、4 列上，只要 A×B 或 C×D 任何一项具有显著性，即使 C 因素对结果影响不大也会被误判为有明显影响，因为它们被安排在同一列上而无法区分。

显然，如果无法否定任何一项交互作用的存在，就只能采用更大的 L12(2¹¹)正交表。

回到表 1，L9(3⁴)正交表共有 4 列，如果不能明确否定 A×B 的存在，由于 A、B 各占 1 列，C、D 只能占其余 2 列，而这 2 列也是 A×B 出现的位置，导致混杂现象；同理，如果不能明确否定 C×D 的存在，由于 C×D 出现的位置正好也是 A、B 所占的 2 列，因此也存在混杂现象，见表 1。因此，若不能明确否定 A×B 的存在，即使仅需考察 A、B 两个因素，使用 L9(3⁴)也不妥(没有位置安排空白列，或空白列只能安排在交互作用列上而产生混杂现象)，必须使用重复测量设计或更大的正交表；如果还要另外考虑 C、D，只能使用更大的正交表而不能依赖重复测量，因为交互作用出现的位置仅与相关主因素的位置有关，而重复测量仅仅解决了误差项的计算问题，并未改变主因素的位置，

也就不能避免混杂现象。

无需参考任何表格,通过 SPSS 也可将各实验因素安排在正确的列上,从而得到一个合理的设计方案。首先利用 SPSS 的正交设计模块生成一个正交设计方案。填入任意的数字作为模拟实验结果,然后在 GLM 中设置你所期待的模型,若能得到有效的方差分析结果就表明设计方案合理,否则改动随机数种子生成另一个新的正交设计方案,然后重新验证。如此反复,直至得出合理的设计方案。

也可以单独使用 GLM 验证自己设计的方案是否合理。例如,假设表 1 只需考虑 A 和 B 两个因素,同时也希望考察 A×B 的显著性,检查 L9(3⁴)的正交设计是否可以满足要求。使用试验号(也可用其他任意的数字)作为 GLM 的因变量(指标值),结果见表 5。

表 5 正交设计合理性检验的方差分析法(SPSS 软件)
Tab 5 Validation of orthogonal design through the results of ANOVA from SPSS software

因素	离均差平方和(SS)	自由度	均方(MS)	F 值	P 值
A	54.000	2.000	27.000	.	.
B	6.000	2.000	3.000	.	.
A×B	0.000	4.000	0.000	.	.
Error	0.000	0	.	.	.

只需检查表 5 各项的自由度。本例误差项(Error)的自由度为 0,而 0 做分母导致了 MS 误差不能被计算[按(2)计算],相应的 F 值和 P 值也就不存在,也就得不出有效的结果,这说明 L9(3⁴)的实验次数不足,需要使用更大的正交表;若某一主因素和某一交互作用的自由度均为 0,说明该主因素与该交互作用存在混杂现象,需要重新调整主因素的列位置,或设置新的随机数种子重新产生正交方案,直至符合要求。

产生表 5 的过程实际上就是一个模拟实验,通过这个模拟实验可以保证正交试验在设计上的可行性,避免试验结束后出现不可补救的错误。

2.3 错误的方差分析结果

国内药学文献中的方差分析多为手工计算,导致了大量的结果错误。这里既有简单的计算错误,也有相当一部分是研究者对方差分析的各种复杂计算公式理解不透而造成的实质性错误。例如,边洪荣等^[10]的方差分析结果中,其所有 F 值

的计算结果全部错误,自然也就导致显著性结果(P 值)错误;钟晓明等^[11]虽然计算结果完全正确,但却无中生有地列出一个误差项的结果(54.19)和自由度(2),导致了原文表 4 的总自由度为 10(原文未累计),超过了真正的总自由度数(总自由度为总试验次数减 1,等于 8),而这是绝对不可能的。事实上,他的误差项就是由因素 D(时间)转化过来的(这个不留空白列的做法也不合理,上面已经讨论过),李小芳等也犯了同类错误^[12];其他计算错误也不罕见^[13-14],仅是错误程度不同而已。

建议采用 SPSS 或 SAS 等专业统计软件中的 GLM 分析这类数据^[15],不但可以保证方差分析结果的准确性,还可以对选择最优工艺提供实质性的帮助(见下文)。

2.4 选择最优工艺不考虑成本因素

文献报道的所有正交试验优选生产工艺的方法基本上不考虑成本因素,这种方法选出的所谓最优工艺很可能从成本-效益的角度上是不划算的,应该也不是生产企业最想要的。

要考虑成本因素,就需要进行各因素生产成本的评估,这过程必须有财务人员的参与,不能单靠实验室研究人员来决定。

假设表 1 的 A 因素代表提取次数,分 3 个水平(1 次、2 次和 3 次),经成本评估确定每增加提取 1 次需要浸膏收率至少增加 1 个百分点才能抵消提取所需的成本(人工、能源消耗等)增加, B 因素为真正的误差项,各因素之间不存在交互作用。在让 GLM 产生表 3 的结果时,还可以在 Options(选项)按钮中设置对主效应进行比较(选择 Sidak 法进行校正),产生表 6 的结果。

表 6 A 因素(提取次数)浸膏收率(%)的主效应比较结果
Tab 6 Pairwise comparisons of the group means(extract rate, %) of factor A

A 因素	A 因素	平均差值	标准误	P 值	90% 置信区间	
I 水平	J 水平	I 水平-J 水平			下限	上限
1 次	2 次	-1.700	0.119	0.014	-2.322	-1.078
	3 次	-2.667	0.119	0.006	-3.289	-2.045
2 次	1 次	1.700	0.119	0.014	1.078	2.322
	3 次	-0.967	0.119	0.044	-1.589	-0.345
3 次	1 次	2.667	0.119	0.006	2.045	3.289
	2 次	0.967	0.119	0.044	0.345	1.589

根据表 6, 提取次数从 1 次提高到 2 次, 浸膏的平均毛收益为 1.700 个百分点(90%置信区间为 1.078~2.322 个百分点), 平均毛收益>1.078 的概率为 95%(95%的单侧界值与 90%的双侧界值相同), 表明在绝大多数情况下能抵消新增成本(1 个百分点), 平均净收益为 0.7 个百分点(90%置信区间为 0.078~1.322 个百分点); 但是, 提取次数从 2 次提高到 3 次, 平均毛收益仅为 0.967 个百分点, 能弥补新增成本的概率不到 50%。因此, A 因素的最优工艺应该为提取 2 次。对这类问题, 如果不考虑成本因素, 单纯从统计学的角度出发选择提取 3 次作为最优工艺将会得不偿失。同理确定 C、D 因素的最优工艺。

3 讨论

目前文献报道的正交试验过于追求减少试验次数(多数不超过 10 次), 从而出现无空白列、无理由忽视交互作用等设计, 这虽然可以节省一些时间和研究经费, 但代价是实验结果不可靠, 得不出真正的最优工艺, 最终得不偿失。事实上, 若实验因素各水平之间确实存在明显差异, 则实验次数的增加将降低表 6 的标准误, 提高检验效能(power), 其潜在效益远大于增加实验次数的成本。

SPSS 的正交设计模块虽然使用方便, 但它不考虑交互作用, 将各因素随机分配到各列中, 常常导致混杂现象, 幸好各个方案对应的种子很多, 一般只需改动几次随机数种子就能得到所需的设计。

对于表 6, 如果 GLM 中没有交互作用项, 也可在的 Post Hoc(后验比较)按钮中设置对主效应进行比较, 在校正方法相同的前提下, 与通过 Options 按钮比较的结果完全相同; 但是, 如果需要考察交互作用项, 只能使用 Options 按钮的比较结果, Post Hoc 的比较结果由于不能修正交互作用对主因素的影响而不够合理。

REFERENCES

[1] ZHAN R R, NI P, LIU D, et al. Chitosan loaded pig thymosin

microspheres for drug deliver [J]. Lett Biotechnol(生物技术通讯), 2012, 23(6): 846-849.

- [2] LI Y J, ZHANG J, HAN H X. Extraction technology of bitter almond seed oil [J]. Xinjiang Med Univ J(新疆医科大学学报), 2012, 35(10): 1344-1347.
- [3] YE Y, OU X H, HUANG Q J. Optimization of forming technology of dihydromyricetin microcapsules by orthogonal experiment [J]. Lishizhen Med Mater Med Res(时珍国医国药), 2012, 23(9): 2251-2252.
- [4] LIN F J. Optimum extraction process of free anthraquinone in Rumex by orthogonal design [J]. J Chin Med Mater(中药材), 2012, 35(9): 1521-1523.
- [5] HE Y, PAN W S. Preparation and *in vitro* release of breviscapin two-layer osmotic pump tablets [J]. West China J Pharm Sci(华西药科学杂志), 2009, 24(1): 25-27.
- [6] JIA S X, WEI Q, LI P, et al. Optimization of prescription of chitosan vaginal bioadhesive film by orthogonal experiment [J]. Chin New Drug J(中国新药杂志), 2008, 17(16): 1422-1424.
- [7] JIANG X H, YU C W, GU H Y, et al. Screening induced conditions of artemisinin multiple shoots by multi-factor orthogonal [J]. China J Chin Mater Med(中国中药杂志), 2008, 33(23): 2854-2855.
- [8] CHEN L, YE L H, ZHU H T, et al. Study on the optimum extraction process of Radix Tinosporae by orthogonal design [J]. Chin Hosp Pharm J(中国医院药学杂志), 2008, 28(23): 2004-2006.
- [9] YU L F, ZHENG S G. Math Stat Meth(数理统计方法) [M]. Shanghai: the Science & Technology Press in Shanghai, 1985: 219-222.
- [10] BIAN H R, SUN G L, ZHANG H L. Study on the optimal ultrasound extraction of Lentinan by orthogonal design [J]. J Chin Med Mater(中药材), 2006, 29(3): 289-291.
- [11] ZHONG X M, HUANG Z, MAO Q Q. Study on the optimal extraction process of Littoralis imperatorin by orthogonal design [J]. Chin Tradit Pat Med(中成药), 2006, 28(5): 749-750.
- [12] LI X F, HUANG Y, BAI L, et al. Optimum process for Kangfuxin drop pill by orthogonal design [J]. Chin Tradit Pat Med(中成药), 2006, 28(12): 1728-1731.
- [13] ZHONG T, CAO C Y, MEI Q X, et al. Optimization of the extraction process of Sihuang paste [J]. China Pharm(中国药房), 2008, 19(12): 908-909.
- [14] HOU J, ZHOU S W. Optimization of the preparation technology of berberine hydrochloride solid lipid nanoparticles by orthogonal experiment [J]. China Pharm(中国药房), 2008, 19(15): 1150-1152.
- [15] MO Z J, CHENG X Y. The application of general linear model to compatible experiments [J]. Chin J Mod Appl Pharm(中国现代应用药理学), 2003, 20(3): 234-237.

收稿日期: 2012-11-21