

# 精神卫生科研如何严格遵守试验设计四原则 之对照原则

沈 宁<sup>1</sup> 胡良平<sup>1 2\*</sup>

(1. 军事医学科学院生物医学统计学咨询中心, 北京 100850;

2. 世界中医药联合会临床科研统计学专业委员会, 北京 100029

\* 通信作者: 胡良平, E-mail: lphu812@sina.com)

**【摘要】** 明确阐释在进行精神卫生临床试验设计时, 应正确把握“对照原则”的意义和要领。从基本常识出发, 并基于精神卫生科研的特点, 寻找和发现在此研究领域中, 怎样做才能被称为严格遵守了“对照原则”。通过结合本专业的特点, 并结合实例, 获得如下的结果: 即在精神卫生临床试验设计时, 必须把握好以下五个方面: ①正确选定合适的对照形式; ②正确领悟什么是合理的对照组; ③必须尽可能设置合理的对照组; ④应尽可能避免出现假对照和对照不全等错误; ⑤应正确确定对照的组数。在如何严格遵守对照原则问题上, 正确把握好前述提及的五个方面, 就是抓住了问题的本质, 是提高临床试验研究质量的一个重要环节。

**【关键词】** 精神卫生; 临床试验设计; 对照原则; 假对照; 对照不全

中图分类号: R195.1

文献标识码: A

doi: 10.11886/j.issn.1007-3256.2016.04.002

## Adherence to four principles in the clinical trial design of mental health: the "control principle"

SHEN Ning<sup>1</sup>, HU Liang-ping<sup>1 2\*</sup>

(1. Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China;

2. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China

\* Corresponding author: HU Liang-ping, E-mail: lphu812@sina.com)

**【Abstract】** To elaborate how to develop and implement best practice of the "control principle" in the clinical trial design of mental health. In the context of discipline of mental health, the question challenges designers is how to adherence to the "control principle", the essence of which includes whether there is a control group and whether the control group is properly set up. By combining the general features with the examples in the discipline of mental health, the five aspects below should be taken into account in the clinical trial design. ①To correctly choose an appropriate type of control. ②To perceive and determine what is a proper control group. ③To make a proper control group to the best of one's ability. ④To avoid the mistakes of a false control group and an insufficient control group. ⑤To correctly determine the quantity of control groups. Thus as to adherence to the "control principle", the essence of which requires designers to put the above five aspects into practice, and it's also a pivotal procedure to achieve high-quality findings of clinical trial research.

**【Key words】** Mental health; Clinical trial design; Control Principle; False control; Control insufficiency

### 1 概 述<sup>[1-3]</sup>

#### 1.1 对照与对照原则的概念

何为对照? 对照就是“对比”之意。在试验研究中, 要想得出什么处理或治疗效果, 必须交代“与谁比较”, 比较至少要在两个事物之间进行, 在统计学上通常是两个组(常称为试验组与对照组)。设立对照组实际上就是为试验组寻找一个参照物或对比的基础。

何为对照原则? 对照原则就是强调在试验研究中必须设置合理的对照组。换句话说, 任何一个试验研究项目, 若仅依据一个试验组得出结论, 则此结

论没有丝毫说服力。

例如, 假定用某药治疗感冒患者 100 例, 结果是治愈了 82 例, 人们并不能认为此药物对于感冒的治愈率为 82%, 更不能得出此药物治疗感冒很有效的结论。因为若换另外一批感冒患者, 其治愈率不一定也是 82%, 况且, 有些感冒患者还可以自愈。若自愈率为 80%, 则表明被试药物治疗感冒的疗效并不高; 若感冒患者的自愈率仅为 40%, 而且, 两组(药物治疗组与安慰剂对照组)患者身体健康状况和感冒的严重程度等方面都非常接近且例数都在 100 例左右或更多, 通过统计分析, 很可能可以得出该药物治疗感冒的效果较好的结论。这里的“自愈

率”意味着什么?意味着设置了一个“空白对照组”。也就是说,“自愈率”实际上就是测自不服用任何感冒药的感冒患者(他们形成一个空白对照组)的自然康复率。由此可知,在试验研究中,对照原则是不可违背的。不仅应当有对照组,而且,必须设置合理的对照组。对照组的“合理性”,不是一两句话能阐述清楚的,详见本文第 3 部分。

## 1.2 对照原则的作用

对照原则的作用是提高试验的鉴别能力,增强结论的说服力。有了对照组,可将影响因素给受试对象带来的效应(症状、体征或指标的改变)与其他因素(如疾病的自然进展、观察者或患者的期望、其他治疗措施等非试验因素)造成的效应区分开来。出自对照组的试验效应可以告诉人们,在没有接受研究者关注的那种特定处理的条件下,受试对象身上发生了什么情况。“孤掌难鸣”和“王婆卖瓜,自卖自夸”等俗语都清楚地说明一个道理:缺乏对照的研究是没有说服力的。

例如,为探讨某种新药治疗急性脑梗死的临床效果,若只对一组该病患者给予该药物治疗,评价用药后的疗效,虽然有时或许也能观察到患者病情有所好转,但因为没有与其他现有的药物作比较,不能判断是其他非试验因素产生的效应还是此新药产生的疗效,就很难得出这种药物疗效优劣的结论。如果研究者将纳入研究的足够多的患者随机分为两组(两组患者在病情、病程、年龄、性别构成、辅助治疗等方面非常均衡),一组给予试验药物,另一组给予现有的常规治疗药物,则可以鉴别出该试验药物的真实效应。

## 2 对照形式的种类<sup>[1-3]</sup>

### 2.1 安慰剂对照

它是指外观与受试药物相同,且无药理活性的物质。在临床研究中安慰剂对照常用来代替受试药物,以排除精神心理等非药物因素的影响。

### 2.2 空白对照

它是指对照组不接受任何真实的处理。空白对照简单易行,但在以人为受试对象的研究中可能会出现下列两种情形之一:①因为设置空白对照,在某些实际临床问题中违反了医学伦理道德,将导致拟开展的临床试验研究项目不能获得有资质的伦理委员会的批准;②有时并不违反医学伦理道德,但未采

取有效的“盲法”或“遮蔽措施”,容易引起对照组与试验组中的受试对象在心理和依从性上的差异,从而影响试验结果的可靠性。

### 2.3 试验对照

它是指对对照组不施加与试验组相同的处理,但施加试验组也具有相同基础处理。由于试验结果往往会受到试验因素和其他非试验因素影响,设立试验对照可以用来排除某种特定的基础处理所产生效应的干扰。

例如,为了评价一种中草药烟熏剂对空气中某些细菌的杀菌效果,此时的处理就是“带药性的烟雾”。那么,人们不禁要问,“杂草或木柴燃烧后生成的烟雾”是否也有杀菌效果呢?为了打消人们的疑虑,开展此项试验研究时,就必须采用无药物性能的柴草点燃后产生的烟雾作为试验对照,以抵消“烟雾”本身对杀菌效果造成的影响。

又例如,某研究者拟证明儿童每天上午和下午在课间各吃一个含赖氨酸的面包有助于身体生长发育,把“吃含赖氨酸面包”的儿童称为“试验组”,研究者还必须设置一组“吃不含赖氨酸面包”的儿童作为“试验对照组”。这两组儿童,除了所吃的面包中“含”与“不含”赖氨酸不同外,在其他一切方面(面包的大小、颜色、口味等都完全一样;在试验期间儿童在同一所学校住宿、作息时间相同、饮食结构和量接近、学习负担和身体锻炼情况接近、不接受儿童家长给予的任何物质和精神帮助)应尽可能一致。

### 2.4 自身对照

对照与试验在同一受试对象身上进行。“对照”应视为接受“处理 1”,而“试验”应视为接受“处理 2”。在实践中,常会出现下面两种情形,第二种情形要慎用。

第一种情形“处理 1”为空白(即不给予处理)、“处理 2”为真实处理。

例如,假定有一组从未服用过降血压药物且已被确诊为高血压患者,在给他们服用一种特定剂量的降血压药物之前与之后(假定服药后半小时)分别测定每位患者的收缩压数值。将每位患者在这两个时间点上测得的收缩压数据配成一对,就叫做“自身配对设计”的一元定量资料。这就是最常见的“自身对照”临床应用的实例。在这个实例里,第 1 次测定收缩压数值时,实际上未服用降血压的药物,而第 2 次测定收缩压数值时,已经服用降血压药

物半个小时了(在本质上,这是药物及其作用时间在共同起作用的时间点,“服药后何时测定收缩压数值的时间长度”应基于临床经验或预试验的结果来确定,应是此药物起效的最佳时间)。

注意:第 1 次测定与第 2 次测定之间的时间间隔不应过长。因为一旦时间间隔过长,随着时间的推移,指标的数值本身会发生明显的改变,其效应会与“处理 2”产生的效应叠加在一起,使得第二次测定的结果不能真实地反映单纯“处理 2”的效应。

第二种情形“处理 1”为第一种真实处理、“处理 2”为第二种真实处理。

例如,假定有一组从未服用过降血压药物且已被确诊为高血压的患者,先给他们服用降压药 A,半小时后测定收缩压数值;过半小时后,再给他们服用降压药 B,半小时后再测定收缩压数值。这种“自身对照”设计是不科学的,因为第 2 次测定的收缩压数值受到了第 1 次使用的 A 药的影响。即使时间间隔再长一些,也不妥。

注意:若可从每个个体身上取出样品(如抽取血液样品)均分成两份,分别施加不同的药物或处理,这种自身对照结果的准确程度就比上面提及的“第一种情形”更高一些,因为它几乎不存在“时间(两次测定指标数值的时间间隔)”的效应了。

若是“活体或整体试验”,两次都是“真实处理”,与其对应的“试验因素”最好不是“药物种类”或“剂量大小(不含零剂量)”;若是类似“体温计种类”、“测重仪种类”这样的“试验因素”,其结果的准确度就很高;若“试验因素”是“血压计种类”,当其用于“自身对照”时,其准确程度介于“药物种类或剂量大小(不含零剂量)”与“体温计种类或测重仪种类”之间。

由此可知,在拟选用自身对照设计时,一定要注意“试验因素”究竟是什么,关键是要考察它的任何一个水平是否会改变观测指标的取值。若会改变如“药物种类”或“剂量大小(不含零剂量)”,应尽可能不采用自身对照形式。

## 2.5 标准对照或设置金标准

用现有标准方法、常规方法、标准值或参考值作为对照。试验研究中常用于某种新方法是否可代替传统方法的研究。

例如,在临床上,常用“病理检查法”作为标准对照,而用新研究出来的“检查方法”作为“试验组”。这是考察“两种检测方法”检测结果是否一致时,常用的一种对照形式。

## 2.6 中外对照与历史对照

中外对照是将中国做的某个试验作为“试验组”,而将外国做的相同试验作为“对照组”的研究。历史对照是将研究者以往的研究结果或文献上他人的研究结果与本次研究结果作对照。这两种对照通常意味着试验组与对照组在“受试对象、环境、条件、时间、研究者、测定方法和评价方法等”很多方面不一致,“对照组”与“试验组”之间的可比性差,说服力不强,应尽量少用或不用。

## 3 设置合理对照组的关键技术<sup>[1-3]</sup>

### 3.1 何为合理的对照组

在单因素试验研究场合下,合理的对照组应当是与其他各组最具有可比性且从伦理学等方面考量都符合要求的那个组。例如,为了通过临床试验回答某种治疗妇女更年期综合征药物的效果与安全性,假定此药物只设置一个特定的剂量,则试验组就是用该药的特定剂量组,那么,合理的对照组究竟是什么呢?若使用安慰剂不违反医学伦理道德,则使用“安慰剂组”就是合理的对照组;当然,使用目前国际上疗效最好、安全性最高的某“阳性药”也是合理的对照组。

前面仅仅是从“试验因素的不同水平”角度来谈论。用“安慰剂”作为对照药时,试验因素的名称可以叫做“药物种类”,其两个水平分别叫做“新药”与“安慰剂”;但试验因素的更确切名称为“新药用否”,其两个水平分别叫做“用新药”与“不用新药”。而用“阳性药”(注意,“阳性药”的剂量应当是其疗效最好且安全性又很高的剂量)时,试验因素的名称只能叫做“药物种类”,其两个水平分别叫做“新药”与“阳性药”。事实上,一个对照组究竟能否被称为“合理的对照组”,除了前面述及的“试验因素的水平”确定得是否合理以外,还应考察对照组中的受试对象在一切非试验因素的取值上是否尽可能与试验组中的受试对象一致。所以,一个对照组只有在前面提及的两个方面都符合要求时,才能被称为“合理的对照组”。

在多因素试验研究场合下,怎样的对照组才能被称为“合理的对照组”呢?这是一个很难用简单的语言说清楚的问题。一般来说,应从多个试验因素中,将每个独立的试验因素抽提出来,在每个试验因素的全部水平组中,确定一个水平组作为其“合理的对照组”。有时,将某一个试验因素的全部水

平组视为“互为对照组”,更符合人们的认识水平或实际情况。例如,假定“某药物的剂量”是多个试验因素中的一个试验因素,其四个水平分别是“2、4、6、8 mg”。此时,将这四个剂量组视为“互为对照组”更合情合理。

综上所述,无论在单因素还是多因素的试验研究场合下,千万不要机械地去规定哪个组就一定是“对照组”,其他组就一定是“试验组”。关键是所做的安排或设置的全部组,无论从什么角度去挑毛病,都挑不出“硬伤”,那么这个“设计”就是科学的、严谨的。设置“合理的对照组”只是科学完善科研设计方案中的一个“细微环节”,大目标实现了即可,不必“小题大作”。

为了使实际工作者在设置对照组时便于操作,下面总结出设置合理对照应满足的一些要求。

### 3.2 合理对照组应满足的基本要求

#### 3.2.1 应满足均衡性要求

通过随机化、盲法等手段,使试验组与对照组除所考察的试验因素的水平取值不同外,其他重要的、可控制的、对试验结果有影响的试验因素和非试验因素所起的作用应尽量一致。如在动物试验中,动物的来源、种属、性别、体重、健康状况应尽量相近。在以人为受试对象时,受试对象的年龄、性别、病情、种族等基线资料应当相同或相近。

#### 3.2.2 应遵循专设、同步的要求

所谓专设,是指在临床试验中,将合格的受试对象分出部分受试对象作为对照,在试验结束时比较两组的处理效应才能达到对照组所起的“比较鉴别”作用。一般情况下,不宜只观察试验组的结果,将该结果与研究本人或他人过去的研究结果进行比较。

所谓同步,就是要求设立平行对照组,从与试验组相同的人群中选择对照,同时按照各自规定的方法进行治疗。

#### 3.2.3 对照形式应与研究目的和内容相符的要求

例如在药物临床试验中,尽量选择安慰剂对照或标准对照或阳性药对照,而不选择空白对照,只有当安慰剂盲法试验无法执行,或执行起来非常困难,又没有合适的标准对照或阳性药对照,且不会严重违背伦理道德时,才选择空白对照。

#### 3.2.4 对照组应是因素的某个称职水平的要求

若试验安排中只有一个合理的对照组,那它必

须与其他组之间形成一种高度可比的关系,即全部组共同形成一个“因素”,其中的对照组一定是该因素的一个称职的水平。也就是说,在只有一个合理的对照组的安排中,它一定是一个标准的单因素设计类型;若试验安排中容纳了  $m$  个合理的对照组,那这个试验安排中很可能就有  $m$  个因素。

## 4 对照组的数目

### 4.1 问题的提出

很多学过应用统计学(特别是医学统计学、卫生统计学、临床统计学和生物统计学)的人脑海中似乎都有一个很深的印象:在一个试验研究问题中,必须设立一个对照组。因此,无论该试验研究项目实际上涉及多少个影响因素(包括试验因素和重要非试验因素),制订试验设计方案的人一般都只设立“一个对照组”。问题是这样的思维观念和做法何时是正确的、何时又是错误的呢?

### 4.2 只设一个对照组是否总是合理的

试验设计中只设置一个对照组是否总是合理的?应视具体情况而定。若研究者的试验中仅涉及一个试验因素,一个对照组已足够。例如,研究一种药物的一个特定剂量的疗效,只要设置一个空白对照或标准对照或安慰剂对照。若研究者的试验涉及到多个因素,为了能正确评价因素间的交互作用效应大小,往往要求每个试验因素都有相对应的对照组。否则,如果对照设置不全(即有些组没有与其对应的对照组),某些试验因素就有可能成为其他试验因素的“混杂因素”了,从而破坏了那些试验因素组间的均衡性。下面的两个实例清楚地说明了这一点。

【例1】某科研工作者做了一项试验研究,其试验安排见表1。试问:该研究者仅设置一个空白对照组是否合适?

表1 一个涉及手术与用药的多因素试验安排

组别	处理方法
A组( $n=8$ )	去势+结扎+用药
B组( $n=8$ )	去势+结扎+不用药
C组( $n=8$ )	结扎+用药
D组( $n=8$ )	结扎+不用药
E组( $n=8$ )	去势+不用药
F组( $n=5$ )	空白对照组(即假手术组)

【解说】在表1中,表面上只有一个对照组,即

“F 组(空白对照组)”或“F 组(假手术对照组)”。但这个对照组有什么意义呢?事实上,这个“F 组”仅仅是“D 组”与“E 组”的“合理对照组”,而没有资格成为其他组的“对照组”。因为“D、E、F”三组中的动物都没有服药,只是接受的手术方式不同而已,因此,“手术方式”是这三组的试验因素,其水平分别为“结扎(D 组)”、“去势(E 组)”和“假手术(F 组)”。在本例中,仅设立一个对照组出错的原因在于此项试验研究不是单因素多水平设计,而是涉及到三个试验因素(即“去势与否”、“结扎与否”和“用药与否”)的一种不正确的安排(因为三个因素水平组合不完全),故它没有资格被称为一种标准的多因素设计。

【例 2】某科研工作者做了一项试验研究,其试验安排见表 2。试问:该研究者仅设置一个试验对照组是否合适?

表 2 一个关于氧吸服试验的多因素安排

组 别	处理方法
对照组( $n=20$ )	吸压缩空气,每天晚上一次,每次 60 min,连续 3 天
吸氧 1 组( $n=20$ )	每天早、中、晚吸氧一次,每次 60 min,连续 3 天
吸氧 2 组( $n=20$ )	每天中、晚吸氧一次,每次 60 min,连续 3 天
吸氧 3 组( $n=20$ )	每天晚上吸氧一次,每次 60 min,连续 3 天
吸氧 4 组( $n=20$ )	每天中午、晚上吸氧一次,每次 60 min,连续 7 天
吸氧 5 组( $n=20$ )	每天晚上吸氧一次,每次 60 min,连续 7 天

【解说】在表 2 中,研究者心理盘算:用第一行上的“吸压缩空气的对照组”为基准,从而说明后面的 5 个试验组中,哪一个效果最好。这是把整个试验视为“单因素六水平设计”了,问题是这个因素的名字叫什么?显然,只有“组别”或“处理”或“方案”这样含糊其辞或十分笼统的词组才能表达所有组的“含义”。这说明此试验安排是一个“假单因素多水平设计类型”,这个“对照组”仅仅是“吸氧 3 组”的合理对照组,而并非其他“吸氧组”的对照组。因此,仅设置这一个“对照组”是不合格的。

值得一提的是,在上面两个实例中,仅设立一个“对照组”出错的根本原因在于:试验研究项目本身不是一个标准的单因素多水平设计类型。在本质上,涉及到多个试验因素,而在安排中,既没有将全部因素的水平进行全面组合,又没有数学根据地抽取其中的部分组合,从而导致一个“拍着脑袋定方案、随心所欲搞科研”的“产物”出现。

#### 4.3 多种对照形式并存

在一个试验研究项目中“只设立一个对照组”的观念和认识在多因素试验研究场合中必须坚决被摒弃!事实上,有时在同一项试验研究中,不仅需要设立多个对照组,而且还同时需要运用多种对照形式,如安慰剂对照和试验对照同时并存、试验对照与阳性对照并存。其实,安慰剂对照很可能是某个因素的一个水平;而试验对照很可能是另一个因素的一个水平。由于人们的习惯思维,喜欢指定某一个或两个组为对照组。事实上,在多因素的试验场合下,一般来说,同一个因素的各水平组之间形成“相互对照”,只是人们不太习惯这样认识而已。

##### 4.3.1 多因素水平组合完全且平衡的安排

【例 3】某研究者进行绵马贯众及单芽狗脊贯众饮片的凝血时间对比试验,取昆明种小鼠 80 只,雌雄各半,按体重、性别随机分成 8 组,分别为绵马贯众生品高低剂量组、炭品高低剂量组,单芽狗脊贯众生品高低剂量组、炭品高低剂量组。各组小鼠连续灌胃给药 3 d,第 3 天给药 1 h 后以毛细管法测定小鼠凝血时间。给药剂量及凝血时间见表 3。试问:这是单因素八水平设计定量资料吗?

表 3 药物对凝血时间的影响( $\bar{x} \pm s$ )

组 别	剂量 (g 生药/kg 体重)	凝血 时间
绵马贯众生品高剂量组( $n=10$ )	2.25	155.5 ± 28.2
绵马贯众生品低剂量组( $n=10$ )	0.75	164.2 ± 25.7
绵马贯众炭品高剂量组( $n=10$ )	2.25	128.6 ± 23.8
绵马贯众炭品低剂量组( $n=10$ )	0.75	141.0 ± 23.0
单芽狗脊贯众生品高剂量组( $n=10$ )	2.25	158.2 ± 27.0
单芽狗脊贯众生品低剂量组( $n=10$ )	0.75	169.4 ± 26.8
单芽狗脊贯众炭品高剂量组( $n=10$ )	2.25	124.3 ± 29.5
单芽狗脊贯众炭品低剂量组( $n=10$ )	0.75	138.2 ± 16.6

【解说】表 3 中“组别”下共有 8 个水平,很多科研人员会想当然地认为这个定量资料就是单因素八水平设计定量资料。其实,这样理解是错误的,他们被“假象”蒙住了眼睛。“组别”在这里是一种“复合型因素”的替身,它实质上是三个两水平因素的全面组合,即“药材”、“剂量”和“炮制类型”。表 3 是该定量资料的“原型”(将问题本质揭示出来的数据结构),其标准型(可以直接进行统计分析的数据结构)见表 4。

表 4 药物对凝血时间的影响 (n = 10)

药材种类	炮制类型	凝血时间( $\bar{x} \pm s, s$ )	
		剂量( g/kg):	
		2.25	0.75
绵马贯众	生品	155.5 ± 28.2	164.2 ± 25.7
绵马贯众	炭品	128.6 ± 23.8	141.0 ± 23.0
单芽狗脊贯众	生品	158.2 ± 27.0	169.4 ± 26.8
单芽狗脊贯众	炭品	124.3 ± 29.5	138.2 ± 16.6

若有专业依据,认定三个试验因素对观察结果的影响有主次之分时,本定量资料应属嵌套设计定量资料;若专业上无法判断三个试验因素对观察结果的影响是否存在主次之分时,本定量资料应属三因素析因设计定量资料。

“绵马贯众”与“单芽狗脊贯众”之间形成“相互对照”;“生品”与“炭品”之间形成“相互对照”;而“2.25”与“0.75”(g 生药/kg 体重)两种剂量之间形成“相互对照”。也就是说,只有在同一个试验因素的不同水平之间,才能谈论哪个水平是“对照水

平”,其他水平是“试验水平”。更确切的表述是:同一试验因素全部水平之间形成“相互对照”的关系。

4.3.2 多因素水平组合不完全但平衡的安排

【例 4】为优化提取板蓝根的工艺条件,取 20 g 板蓝根饮片,采用正交设计试验方法筛选板蓝根中有效成分总有机酸的最佳提取工艺,考察的因素有提取溶剂浓度、溶剂用量、提取次数、提取时间,因素及其水平数见表 5。选用 L9(3<sup>4</sup>) 正交表安排试验,见表 6。试问:在此试验设计中,有几个对照组?

表 5 总有机酸提取工艺影响因素水平

水平	水平取值			
	因素: 乙醇浓度 (A, %)	溶剂用量 (B, 倍)	提取次数 (C, 次)	提取时间 (D, h)
1	50	5	1	0.5
2	70	7	2	1.0
3	90	10	3	1.5

表 6 总有机酸提取正交试验结果

试验号	水平代码				总有机酸含量 (%)	
	因素:	A	B	C		D
1		1	1	1	1	2.82
2		1	2	2	2	4.62
3		1	3	3	3	5.22
4		2	1	2	3	4.67
5		2	2	3	1	4.84
6		2	3	1	2	3.81
7		3	1	3	2	3.38
8		3	2	1	3	2.21
9		3	3	2	1	2.65
K <sub>1</sub>		12.66	10.88	8.84	10.31	
K <sub>2</sub>		13.32	11.77	11.94	10.81	
K <sub>3</sub>		8.04	11.68	13.44	12.10	
R		5.28	0.89	4.60	1.79	

【解说】在表 6 中,反映出 4 个因素的 9 种水平组合(即表身中的 9 行),由于每个因素有 3 个水平,它们水平的全面组合应该有 81 种,现在实际选用了其中的 1/9 种组合。表 6 中的 9 行所决定的试验条件被称为“9 个试验点”,虽然因素的水平组合不完全,但这 9 个试验点是依据“正交原理”挑选出

来的,这 9 个试验点在空间具有“均匀分散”性;在统计分析时具有“整齐可比”性。这就是为什么在因素之间交互作用不复杂时,用正交表挑选出来的少数“试验点”能比较好地达到或接近“全部试验点”的试验效果的奥秘之所在。

在表 6 中,有人可能认为:第一行上四个因素都

取“1 水平”,用此行对应的条件作为“对照组”,即以低水平组合为对照组。其实,真实情况并非如此。

比较第一行与第二行不难发现,因素 A 均取“1 水平”,而因素 B、C、D 分别是“1 水平”与“2 水平”之间的比较。

比较第一行与第三行不难发现,因素 A 均取“1 水平”,而因素 B、C、D 分别是“1 水平”与“3 水平”之间的比较。

比较第一行与第四行不难发现,因素 B 均取“1 水平”,而因素 A、C 分别是“1 水平”与“2 水平”之间的比较,因素 D 是“1 水平”与“3 水平”之间的比较。

第一行与其他行之间的关系就不逐一细说了。总之,不能将第一行视为“对照组”。比较合适的说法是:在多因素试验设计场合下,每个因素都以其自身的全部水平进行“相互对照”。没有必要也不应该指定某一种水平组合为其他水平组合的“对照组”。

#### 4.3.3 多因素水平组合不完全且不平衡的安排

在上面的表 1 中,显然,试验涉及到“结扎与否”、“去势与否”和“用药与否”这三个因素,但在表 1 中的安排不是一个某种标准的三因素设计类型,而只能被冠以一项“多因素非平衡组合试验”。正确的做法是把三个两水平因素的水平进行全面组合,共有 8 种不同的水平组合,也就是原试验研究者少安排了两组试验,它们分别是“去势+用药”与“假手术+用药”。全部 8 个组的试验安排可被称为“三因素析因设计”结构或“ $2 \times 2 \times 2$  析因设计”结构。

为什么要加上“结构”两个字而不直接说成“三因素析因设计”或“ $2 \times 2 \times 2$  析因设计”呢?那是因为标准的“析因设计”还需满足其他条件,此处不便赘述。

在上面的表 2 中,情况就复杂得多了。从表面上看,这个安排中似乎也涉及到三个因素,即“吸入物种类(氧气与压缩空气)”、“每天吸的次数(1、2、3 次)”和“疗程(3 d 与 7 d)”。这三个因素水平的全面组合似乎只有  $2 \times 3 \times 2 = 12$  种,但这显然不符合实际情况。因为“每天吸的次数(1、2 和 3 次)”是一个“复合型因素”,其中,每天吸一次又有 3 个水平,即早、中、晚;而每天吸两次也有 3 个水平,即早与中、早与晚、中与晚。

如何将表 2 中的试验安排修改成某种标准的多因素试验设计类型,需要处置好那个复合型的因素。正确答案留给读者去寻找,此处从略。

## 5 对设置对照组方面存在的不足进行辨析与释疑

**【例 5】**两组均给予舍曲林(左洛复,辉瑞公司)治疗,初始剂量 50 mg/d,1 周后加至 100 mg/d,根据疗效及耐受情况,调整剂量至 50~200 mg/d,其中研究组平均剂量为  $(135.84 \pm 12.65)$  mg/d,对照组平均剂量为  $(139.36 \pm 9.46)$  mg/d,两组舍曲林平均剂量比较差异无统计学意义( $P > 0.05$ );研究组舍曲林联合电针治疗,使用电针仪(电量调至患者感到舒适而穴位局部皮肤肌肉轻微抽动为度,使用疏密波,频率为 80~100 次/min,每次 30 min),穴位选取:百会、印堂、内关、足三里,电针治疗每日 1 次,每周 5 次,两组均治疗 4 周<sup>[4]</sup>。

**【辨析与释疑】**本例的对照组可以近似视为“标准对照”,即使用舍曲林治疗创伤后应激障碍。研究组是在用舍曲林治疗的基础上,加用电针治疗。问题是:所选用的“标准对照”是否是最合理的对照?关键在于弄清研究组比对照组“多了什么”。表面上看,“多了电针刺激”。但事实上,还多了“选取的穴位”和“患者心理期待”。所以,比较好的对照形式应当是“试验对照”,即对照组应在使用舍曲林治疗的基础上,也在“百会、印堂、内关、足三里”4 个穴位上采用“假针刺”,以最大限度地降低对照组中受试对象对于“针刺”操作可能带来的“好疗效”的“安慰剂效应”。从而真正显露出“舍曲林联合电针疗法”与“单纯舍曲林治疗”疗效之间的差别大小。

文献[5]在对照组的设立上,也存在与本文例 5 同样的问题,因篇幅所限,就不展开赘述了;文献[6]在对照组的设立上,采取了“假刺激”,从而消除了本文例 5 中存在的弊端,故可认为属于设立了“合理的对照组”。

**【例 6】**改良电痉挛治疗(MECT)组与药物组:为 2014 年 2 月-7 月在解放军 256 医院住院的精神分裂症患者(入组标准与排除标准此处从略)。研究期间所选患者行 MECT 治疗的作为 MECT 组,共入组 74 人,脱落 16 人(其他细节从略);未行 MECT 治疗的作为药物组,共入组 53 人,脱落 12 人(其他细节从略);对照组:为同一时期医院健康工作人员(入组标准和排除标准此处从略),共计入组 82 人。观测的定量指标有 IL-1(pg/mL)、IL-4(pg/mL)、IL-6(pg/mL)、IL-10(pg/mL)和 CRP(mg/L)。对 MECT 组患者在 4 个时间点(首次治疗

前、首次治疗后、4 次治疗后、6 次治疗后) 对上述定量指标进行观测、对药物组患者在 2 个时间点(用 药前、用药 2 周后) 对上述定量指标进行观测、对健

康对照组在 1 个时间点对上述定量指标进行观 测<sup>[7]</sup>,见表 7。试问: 在此项临床试验研究中,对照 组设立得是否合理?

表 7 各组间及 MECT 治疗前后血清 IL-1、IL-4、IL-6、IL-10、CRP 水平比较( $\bar{x} \pm s$ )

组 别	时 间	IL-1( pg/mL)	IL-4( pg/mL)	IL-6( pg/mL)	IL-10( pg/mL)	CRP( mg/L)
MECT 组 (n=58)	首次治疗前	91.97 ± 234.23	25.05 ± 85.52	52.33 ± 173.40	48.82 ± 154.80	1.75 ± 3.60
	首次治疗后	131.40 ± 352.70	45.39 ± 163.73	55.43 ± 169.12	49.13 ± 154.67	1.98 ± 3.65
	4 次治疗后	113.32 ± 312.96	32.07 ± 94.85	61.27 ± 210.69	70.44 ± 235.07	4.29 ± 7.05 <sup>ab</sup>
	6 次治疗后	116.56 ± 259.32	31.73 ± 79.79	45.86 ± 139.73	46.09 ± 130.39	3.59 ± 0.78 <sup>a</sup>
药物组 (n=41)	用药前	108.74 ± 325.32	36.40 ± 93.35	54.62 ± 198.42	53.98 ± 194.95	2.91 ± 6.72
	用药 2 周后	129.48 ± 348.15	54.91 ± 158.92	42.16 ± 111.40	35.51 ± 83.38	3.95 ± 0.67 <sup>a</sup>
健康对照组 (n=82)		87.10 ± 93.06	19.27 ± 34.22	14.11 ± 35.09	29.40 ± 52.47	0.92 ± 1.60

注: 三组间比较 <sup>a</sup>P < 0.05; MECT 组前后比较 <sup>b</sup>P < 0.05

【辨析与释疑】由表 7 可看出,文献[7]在对照组的设立上,存在的问题比较多。在该文中,设立了 3 个组,其中 MECT 组在 4 个不同时间点上对定量指标进行了观测、药物组在两个不同时间点上对定量指标进行了观测、健康对照组仅在 1 个时间点上对定量指标进行了观测。对每个定量指标而言,相当于有“7 个组”之间的对比关系。

第一,应当明确:将“7 个组”同时考察是没有意义的,因为它不是一个具有 7 个水平的试验因素,它属于“多因素非平衡组合试验”,必须结合基本常识和专业常识将“7 个组”拆分开,再重新组合(如何拆分和如何重新组合,因篇幅所限,此处从略)后进行比较,才有意义。

第二,设立的“健康对照组”似乎意义并不大。假定表 7 中所观测的 5 个定量指标是区分精神分裂症患者与健康人的特异性指标,则健康人与此病患者在这些指标上的取值肯定会有明显差别,故“健康对照组”起不到“标准对照”的作用。

第三,“药物组”只能被视为“标准对照组”,但最好是“试验对照组”。也就是说,应在现在的“药物组”基础上加上“未改良电痉挛治疗”。只有设立了这样的“试验对照组”,才能与这篇论文希望达到的试验研究目的相吻合。当然,要符合伦理学要求。

第四,比较合适的做法是设计如下的三组:“单纯药物组”、“药物联合未改良电痉挛治疗组”和“药

物联合改良电痉挛治疗组”,并且各组都在“首次治疗前、首次治疗后、4 次治疗后、6 次治疗后”4 个时间点上观测定量指标的取值。还应对资料采取正确的统计分析(即带有一个协变量且具有一个重复测量因素的两因素设计一元或五元协方差分析)处理,才能得到令人信服的结论(说明:以“首次治疗前”为协变量)。

## 参考文献

- [1] 胡良平,关雪.如何正确执行对照原则[J].中华脑血管病杂志(电子版) 2011 5(2):43-48.
- [2] 毛玮,胡良平,王琪.试验设计必须遵循对照原则[J].药学服务与研究 2010 10(4):252-255.
- [3] 胡良平.科研设计与统计分析[M].北京:军事医学科学出版社,2012:136-205.
- [4] 周平,张瑶,谭庆荣.电针联合舍曲林治疗创伤后应激障碍的效果观察[J].四川精神卫生 2015 28(6):504-506.
- [5] 金毅琼.帕罗西汀联合重复经颅磁刺激治疗女性更年期抑郁症的对照研究[J].四川精神卫生 2015 28(6):515-518.
- [6] 郭新宇,杨媛,田丽.草酸艾司西酞普兰联合重复经颅磁刺激对改善难治性抑郁症患者执行功能的疗效研究[J].四川精神卫生 2016 29(1):26-30.
- [7] 郭永芳,周小东,付华斌,等.改良电痉挛治疗对精神分裂症患者血清细胞因子及 C-反应蛋白水平的影响[J].四川精神卫生,2015 28(2):119-122.

(收稿日期:2016-07-26)

(本文编辑:吴俊林)