

# 正确把握精神卫生临床试验设计三要素的要领(Ⅲ) ——观测指标

谷恒明,胡良平

(军事医学科学院生物医学统计学咨询中心,北京 100850)

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,正确把握“观测指标”的意义和要领。从基本常识出发,基于精神卫生科研的特点,寻找和发现在此研究领域中,怎样做才能被称为正确把握了“观测指标”。通过结合本专业的特点,并结合实例,获得如下的结果,即在进行精神卫生临床试验设计时,必须把握好以下五个方面,①观测指标的分类;②观测指标的选定;③观测指标取值的测定方法;④观测指标测定时间的确定;⑤观测指标测定次数的确定。在如何确定观测指标问题上,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;观测指标;灵敏度;特异度

中图分类号:R195.1

文献标识码:A

doi:10.11886/j.issn.1007-3256.2016.03.004

## To correctly hold the essentials of "Three Key Points" of the clinical trial design of mental health:Part Ⅲ the "outcomes"

GU Heng-ming, HU Liang-ping\*

(Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China)

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

**【Abstract】** To explain clearly that the meanings and essentials of "outcomes" should be grasped correctly, when conducting the clinical trial design of mental health. Starting from the basic common sense, and based on the characteristics of mental health study, in what way it can be considered to correctly grasp the "outcomes" by searching and finding an approach to this research field. By combining the features with some real examples in the specialty of mental health, the following results acquired: Need to master the following five aspects when conducting the clinical trial design of mental health. ①The classification of outcomes; ②The selection of outcomes; ③The method for measuring the values of outcomes; ④The determination of measuring time of outcomes; ⑤The determination of measuring times. It is an important link to improve the quality of clinical trials to correctly hold the previous five aspects, which is to seize the essence of the issue on answering the question of how to determine the outcomes.

**【Key words】** Mental health; The design of clinical trial; Outcomes;Sensitivity; Specificity

## 1 概述

### 1.1 试验效应和观测指标

试验效应就是影响因素作用于受试对象后所产生的结果,它是通过具体的观测指标来体现的,是研究结果的最终体现,也是试验研究的核心内容。换句话说,观测指标是用来反映影响因素作用强弱的重要“尺子”,必须结合影响因素的性质和特点、仪器、试剂和技术水平等多方面综合考虑,找出“特异性强、灵敏度高、准确、可靠”的观测指标,以“客观指标”为主,以“半客观和主观指标”为辅;如果指标选择不当,未能准确反映影响因素的作用,获得的研究结果就缺乏科学性。因此,选择好的指标是关系科学研究成败的重要环节。

### 1.2 精神卫生科研领域中观测指标的特点

在精神卫生学术论文中,观测指标通常有“单一型”和“复合型”两种表象。所谓“单一型”观测指标,是指每个观测指标仅代表一个特定的专业内容或含义,如血清 IL-1、IL-4、IL-6、IL-10、CRP 水平<sup>[1]</sup>就是五个“单一型”观测指标;又如 HRSD 评分<sup>[2]</sup>、PANSS 评分和 WCST 评分<sup>[3]</sup>均来自某种特定量表上多个项目的综合评分,属于“复合型”观测指标。

## 2 观测指标的两种类别

### 2.1 按功能分类

从功能上分类,观测指标可被分为三类:诊断性指标、疗效性指标和安全性指标。

## 2.2 按主次分类

按照主次分类,观测指标可分为主要指标和次要指标。因为无论是诊断性、疗效性和安全性指标都有很多,应突出重点,不能眉毛胡须一把抓。通常每类指标中的主要指标只设定一个,因为主要指标多了,依据它们的取值,经过统计分析,得出自相矛盾的结果时,不便作出结论。

确定一个研究项目各类指标中的主要指标有以下三点好处:其一,重点突出;其二,便于试验前估算样本含量;其三,可以有效避免统计分析结果出现自相矛盾的现象。

## 3 主要与次要指标的确定

### 3.1 主要与次要诊断性指标的确定

需要对受试对象是否患有某种待研究疾病进行认真检查,作出诊断结论。明确诊断是临床试验设计时制订“纳入标准”的重要条件之一。为确定受试对象是否患有某种疾病,诊断的指标往往并非一个,而有多。

结合基本常识、专业知识及预试验的结果,确定一个最有说服力的指标为主要诊断指标,其他为次要诊断指标。最好有“金标准”,金标准即诊断某受试对象患了某病,该患者就一定患了此病,反之亦然。

### 3.2 主要与次要疗效性指标的确定

主要疗效指标又称主要终点,是与试验目的有本质联系的、能确切反映治疗有效性的观察指标。通常主要疗效指标只有 1 个,偶尔会设定 2~3 个。如果存在多个主要疗效指标时,应该在设计方案中考虑控制 I 类错误的方法,如有 5 个主要疗效指标,对它们进行假设检验时,应调整每次假设检验的显著性水平  $\alpha$  的取值,通常  $\alpha = 0.05$ ,而有 5 个主要疗效指标时,对每个主要疗效指标分析时,将其设定为  $\alpha' = 0.05/5 = 0.01$ 。

主要疗效指标应根据试验目的选择易于量化、客观性强、重复性好,并易于在相关研究领域估计的指标。次要疗效指标是指与试验目的相关的辅助性指标,一般来说,次要疗效指标的数目较多。

### 3.3 主要与次要安全性指标的确定

在临床试验中,安全性评价是非常重要的一个方面。在临床试验的早期,这一评价主要是探索性的,且只能发现常见的不良反应;在后期,一般可通

过较大的样本进一步了解药物的安全性。后期的对照试验是一个重要的以无偏倚的方式探索任何新的潜在的药物不良反应的方法。安全性指标也应有主次之分,通常主要安全性指标只有一个,次要安全性指标可以有多个。

药物安全性评价的常用统计指标为不良事件发生率和不良反应发生率。当试验时间较长、有较大的退出治疗比例或死亡比例时,需用生存分析计算累计不良事件发生率。构成安全性评价的资料则主要来源于不良事件的临床表现、实验室检查等。从受试者中收集的安全性指标应尽可能全面,主要包括受试者出现的所有不良事件的类型、发生时间、严重程度、处理措施、持续的时间、转归以及药物剂量与试验药物的关系。所有的安全性指标在评价中均需被重视,其主要分析方法需在研究方案中说明。

## 4 观测指标的选取

### 4.1 选取的一般原则

观测指标的取值反映影响因素作用于受试对象后所产生的试验效应,这些指标包括定性指标和定量指标等。观测指标的选择必须注意它的针对性,即选用的指标必须与研究目的具有本质性联系,且能确切反映影响因素的效应。否则,不能作出正确的判断。

### 4.2 选取指标时应注意的事项

#### 4.2.1 指标的合理性

所谓“合理”就是指所选指标能真实客观地反映出干预措施的临床效应,在专业上能得到合理的解释。

#### 4.2.2 指标的先进性

课题的先进性一般由指标的先进性体现出来,先进指标更能深入地反映所研究的问题,揭示事物的本质。

#### 4.2.3 指标的客观性

在临床试验中应尽量选择客观性指标,即通过测试仪器或工具获得观测结果的指标而避免选用由研究者主观判断观测结果或根据受试对象主诉获取观测结果的主观性指标。若是主观指标,最好将其客观化。

#### 4.2.4 指标的灵敏性和特异性

灵敏性是指所选用的指标对干预措施反应的灵

敏程度,指标的特异性即检测结果的专一性,以便把假阳性结果控制在最低水平。最好选用灵敏度高、特异性好的指标。

【解说】举一个骨科药物临床试验的真实实例:在评价一种治疗骨折药物疗效的临床试验研究中,研究者确定的主要疗效指标为“骨密度”、“骨钙素”、“胫骨缺损处周长”等。显然,选择上述指标来评价药物对骨折的治疗效果既不敏感又不特异,失去了“针对性”,那些指标用来反映患者“骨质疏松状况”比较合适。那么,哪些指标用来反映骨折愈合效果合适呢?根据基本常识和骨科专业知识得知,应选择“骨形态”、“X线片”、“病理检查结果”和“功能恢复情况”等<sup>[4]</sup>。

在精神卫生科研领域中,在选取观测指标方面,是否存在像上面那样“张冠李戴”的现象,需要具有丰富专业知识的学者花很多时间去细读大量的学术论文,才能给出答案。

#### 4.2.5 指标的精确性

指标的精确性包括指标的精密性与准确度。

#### 4.2.6 指标的经济性

指标的经济性是指检测成本应在受试对象所能承受的范围之内。

#### 4.2.7 指标的标准化

指标的标准化是指对指标的采集方法、部位、时间、保存、运输、测定时间、测定方法、测定条件等方面进行严格地规定,应确保不同的操作者或同一操作者不同时间操作,效果基本相同。若不进行标准化,这些与测定结果有关的非试验因素会干扰研究效果,或者成为混杂因素,影响研究结果的准确性。

### 5 观测指标取值的方法

确定反映试验效应的指标以后,还要规定指标观察的常规方法,如观察方法、标准、时间、记录方法及记录格式等。指标的观察或测量应避免偏性,否则会影响结果的比较和分析。有些总体表现(气色、精神状态、营养状态等)、主观症状(疼痛、抑郁、焦虑等)、动物行为学变化(活跃与否、异常行为等)及某些形态变化(如器官外观、动物毛色、红肿、溃烂等),这些指标的判断带有主观性,研究者的心理常偏向于阳性结果,医生常偏向于新疗法组。

如果试验效应的观察带有偏性,会影响结果的比较和分析。为了消除或最大限度地减少这种偏性,

有三种解决途径:一是通过培训来掌握统一评价标准。二是在试验设计时常采用盲法(blind method),单盲法(single blind method):仅受试对象不知道自己被分在哪一组;双盲法(double blind method):受试对象和试验执行者均不知道受试对象被分在哪一组;三盲法(triple blind method):受试对象、试验执行者和统计分析人员三者均不知道受试对象被分在哪一组。双盲法和三盲法在临床试验中应用非常广泛。三是通过一些精心设计的量表,全面度量某些方面的情况,以便取代简单的提问,有利于提高结论的可信度。

### 6 观测指标取值的时间

#### 6.1 一个观测时间点的确定

通常,当各组受试对象接受特定的处理后,需要在一个恰当的时间点上观测指标的取值,问题是如何确定这个“最佳时间点”。这需要结合基本常识、专业知识和必要的预试验结果来粗估计处理因素发挥作用的时间长短,再参考有经验的同行或专家的意见,经过深思熟虑后确定下来。

#### 6.2 多个观测时间点的确定

在实际科研工作中,有时很难找准一个最恰当的观测时间点。比如,坚持锻炼身体有利于人身体健康,用不同方式锻炼多长时间后,观测身体素质好坏合适呢?这里很难给定一个具体的时间段,此时应多选择几个时间点,并且时间点之间的间隔也要慎重考虑后才能确定。有些试验因素产生试验效应很快,此时观测指标取值的时间间隔就应短一些;而有些试验因素产生试验效应很慢,此时观测指标取值的时间间隔就应长一些。具体短到什么程度、长到什么程度,必须结合基本常识、专业知识和预试验结果来综合考虑后再确定。

### 7 对选取观测指标方面的错误辨析与释疑

#### 7.1 指标选取中常见的问题及对策

在临床科研(特别是心理测量、精神卫生研究和中医中药研究等)中,就如何选择观测指标这一环节来看,有如下普遍存在且不容忽视的问题:第一,研究者要么罗列出几十项观测指标,且不告知哪些是主要的疗效性与安全性指标、哪些是次要的疗效性与安全性指标;第二,要么运用几十项、甚至上百项内容的“测量量表”,并采用量表中多项得分之和作为主要评价指标的观测值。

第一种做法的不妥之处在于:在众多指标中分不清主次,导致“眉毛胡须一把抓”和“出现互相矛盾的分析结果”等混乱局面和无法作出肯定结论的尴尬结局;第二种做法的不妥之处在于从表面上看,采用量表总评分反映每位受试者的试验效应似乎是合情合理的,其实不然,因为量表中有些项目是高优指标(即指标取值越大越好),有些项目是低优指标(即指标取值越小越好);另外,众多指标彼此之间的重要性程度不同,如某两项指标均得 8 分,但反映患者病情的实际改善程度可能相差很多。

由于上述两方面原因,对量表中的几十项、甚至上百项内容所得分值简单求和作为主要评价指标的取值,其科学性与合理性难免受到质疑。

比较科学且具有可操作性的做法是:结合基本常识和专业知识,并汇集众多同行专家的意见和建议,从众多的疗效指标中筛选出 1~2 个主要指标,同样,从众多的安全性指标中筛选出 1~2 个主要指标,其他指标一律视为次要指标。若实在无法找出少量的主要指标,可找出 5~10 个相对重要的指标,先对它们进行“趋同化处理”,即将它们都转变成“高优指标(如治愈率)”或“低优指标(如死亡率)”,再采用综合评价方法<sup>[5-6]</sup>将它们合并成一个复合型的“主要指标”。

## 7.2 观测指标的时间点选择不合适

例如,在一个关于某感冒药临床试验的疗效与安全性评价的临床试验资料中,设计者将评价疗效的时间点定在服药后 2 周末。请问:妥否?

【解说】显然,观测指标取值的时间点选得不合适。因为感冒是常见的小毛病,很多人感冒后不接受任何药物治疗,1 周之后就自愈了。服药后 2 周再来观测疗效,试验组与对照组的“疗效”可能非常接近了。即便试验药与对照药疗效明显不同,到 2 周末,自然就把“自愈”的效应也包括进去了。所以,在这个临床试验研究项目中,可能比较恰当的观测时间点为服药后“1 周末”。

下面有一个出给学生做的考题,其中,部分学生给出的答案令人啼笑皆非。

统计学老师曾出过这样一道统计学考题:为了通过试验研究,证明“早睡早起身体好”的古训是否正确,请写出试验设计方案。

【解说】很多考生根据统计学教科书和自己学习试验设计知识后的体会,写出了初步的试验设计方案。其中,在确定观测主要评价指标(假定为肺活量)的时间点这个问题上,相当多的考生给出的答案是“试验开始后 2、4、6、8、10、12、14、16、18、20 小时”观测结果。

这样确定的时间点是没有实用价值的!因为身体素质的提高与否,是一个比较缓慢的过程,不可能锻炼后 2 小时就产生“效应”,时间间隔为 2 小时,也是不切合实际的。

## 8 小 结

试验效应(更具体地说,即观测指标)是试验设计三要素之一,其实,它也是调查研究中的重要要素之一。任何一个没有观测指标的研究项目都是没有实际意义的,然而,即使有了观测指标,研究结果也不一定有意义。本文对与“观测指标”有关的内容作了比较系统的归纳和总结,因篇幅所限,可能还不尽完善。需要时,读者还可参考有关文献<sup>[7-10]</sup>。

## 参考文献

- [1] 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.
- [2] 牛亚杰,杨少杰,杨可冰,等.西酞普兰合并改良电休克治疗对抑郁发作的疗效及认知功能的影响[J].四川精神卫生,2015,28(2):123-126.
- [3] 韩勇,卞清涛.氯氮平联合阿立哌唑对长期住院精神分裂症患者疗效及认知功能的影响[J].四川精神卫生,2015,28(2):131-133.
- [4] 胡良平.实用医学统计学[M].北京:金盾出版社,2014:1-11.
- [5] 胡良平.科研设计与统计分析[M].北京:军事医学科学出版社,2012:469-486.
- [6] 孙振球.医学统计学[M].北京:人民卫生出版社,2002:373-396.
- [7] 胡良平,王琪,毛玮,等.如何正确把握实验设计的三要素[J].中国骨伤,2009,22(4):324-326.
- [8] 王琪,胡良平.试验设计的三要素[J].药学服务与研究,2010,10(2):85-88.
- [9] 胡良平,关雪.科研课题的研究设计与统计分析-如何正确把握试验设计的三要素[J].中华脑血管病杂志(电子版),2010,4(4):46-50.
- [10] 胡良平.面向问题的统计学——(1)科研设计与统计基础[M].北京:人民卫生出版社,2012:42-89.

(收稿日期:2016-02-18)

(本文编辑:唐雪莉)