

## 如何在骨科科研中正确应用对照原则

胡良平, 毛玮, 王琪, 陶丽新

(军事医学科学院生物医学统计学咨询中心, 北京 100850)

Zhongguo Gushang/China J Orthop & Trauma, 2009, 22(7): 560-562 www.zggszz.com

骨科科研中经常采用实验研究, 其目的是验证研究假设是否正确, 得出确切的研究结论。在实验研究中, 往往需要设立对照以便鉴别真伪, 验证研究假设, 因此在实验设计中必须严格遵循对照原则。对照原则在实验研究中非常重要, 如果不遵循该原则或仅形式上遵循, 所得科研结论的可信度就将大打折扣。我们在审稿过程中发现科研论文中经常在对照原则上出问题。本文将从基本常识、专业知识和统计学的角度, 结合《中国骨伤》文章中的具体实例, 探讨在骨科科研中如何把握好对照原则这一问题。

### 1 何为对照原则

所谓对照, 是指在实验研究中, 要想得出什么处理或治疗更佳, 必须交代“与谁比较”, 比较至少应在两个或多个同类事物之间进行, 在统计学上通常是两个或多个组之间进行, 即在实验组与对照组之间比较。设立对照实际上就是寻找一个参照物或对比的基础。对照原则的作用是提高实验的鉴别能力, 增强结论的说服力。

需要强调的是, 对照组的概念还必须予以拓展, 在多因素的实验设计中, 为了考察因素间的交互作用, 往往要求各个实验组都应该有相应的对照组, 否则, 如果对照组设置不全, 就不能很准确地反映某些因素取不同水平对定量指标的影响。因此, 不要以为在任何实验研究中设立了 1、2 个对照组, 就一定能达到研究目的<sup>[1]</sup>。

### 2 设置对照的种类

对照有很多种类型, 可根据研究目的和内容加以选择。常见的对照形式如下。

**2.1 安慰剂对照** 安慰剂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常用用来代替受试药物, 以排除精神心理等非药物因素的影响。

**2.2 空白对照** 即对照组不接受任何处理因素。空白对照简单易行, 但在以人为受试对象的研究中容易引起对照组和实验组在心理上的差异, 从而影响实验结果的可靠性<sup>[2]</sup>。

**2.3 实验对照** 对照组与实验组在基本条件(或基础处理)上完全相同, 这种基本条件可能对观测结果有一定的影响。例如, 对某病患者采取统一的常规治疗, 在此基础上, 将全部患者完全随机均分为两组, 一组不给予其他药物治疗(被称为对照组), 另一组给予 A 药治疗。目的是在常规治疗基础上, 考察 A 药用与不用所对应的疗效之间的差别是否具有统计学意义<sup>[3]</sup>。

**2.4 自身对照** 对照与实验在同一受试对象身上进行。但应

注意两次观测之间的时间间隔不应过长, 否则, 随着时间的推移, 同一个体自身也会发生较大改变, 所产生的效应必然在实验因素不同水平之间产生混杂。

**2.5 标准对照** 用现有标准方法、常规方法作为对照。实验研究中常用于某种新方法是否可代替传统方法(通常被视为当前条件下的标准方法)的研究之中。

**2.6 历史对照** 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。这种对照一般受时间、条件、环境等的影响, 其可比性较差, 结论的可信度很差, 能不采用时尽量不要采用此种类型的对照。

值得一提的是, 有时在同一项实验研究中需要同时运用多种对照形式, 如安慰剂对照和实验对照同时并存, 有时甚至还会出现阴性对照、阳性对照等对照形式。此时, 所对应的实验安排常常不是一种标准的实验设计类型, 而叫做“多因素非平衡组合实验”更为贴切。

### 3 人们在运用对照原则时常犯的错误类型及实例辨析

**3.1 缺乏必要的对照组** 未设置任何对照组。骨科科研中为评价某种疗法的效果, 常常用到回顾性研究, 把该疗法治疗的情况进行总结分析。严格来说, 这种研究并不是一个合格的实验研究。比如下面的例子。

例 1:《Ⅲ度开放性胫腓骨骨折外固定治疗策略》<sup>[4]</sup>一文, 为评价外固定支架的疗效, 对Ⅲ度开放性胫腓骨骨折 120 例 128 个患肢进行了早期外固定支架治疗。结果: 120 例中, 111 例优良, 或通过Ⅱ、Ⅲ期手术, 软组织修复及骨折修复良好; 1 例患者因感染骨折复位不良、不愈合。8 例Ⅱ期截肢。结论: 对于Ⅲ度开放性胫腓骨骨折早期以外固定支架为主的治疗是一种理想的方法。这样的结论可信吗?

辨析与释疑: 该研究只是对一种手术方法进行了评价, 所有受试对象均接受同一种处理, 并没有与目前其他方法进行对比, 所得结论说服力不强, 故此实验在设计上犯了“缺乏必要对照组”的错误。正确的做法是设立合理的对照组, 对照组可以是常规或标准疗法, 采用分层随机方法将全部该病患者均分入两组, 确保两组患者在病情轻重、年龄和性别构成等方面具有很好的均衡性, 选定恰当的评价骨折疗效的指标, 选用合适的统计分析方法比较两组的疗效, 这样得出疗效好坏的结论才比较令人信服。

例 2:《内、外固定加中药治疗浮肘损伤》<sup>[5]</sup>一文, 对 26 例浮肘损伤病例, 采用内外固定加中药治疗, 随访 6 个月~2 年, 疗效按照 Flynn 标准评分: 优 8 例, 良 14 例, 可 3 例, 差 1 例,

总有效率 96.4%。结论是:中西医结合治疗浮肘损伤更能提高治疗浮肘损伤疗效。这样下结论妥当吗?

辨析与释疑:该研究是一种联合治疗方法的评价,用以说明中药对浮肘损伤辅助治疗作用,但是所有受试对象均接受中西医结合治疗,并没有与不用中药进行辅助治疗的情况进行比较,不能验证中药在疗效中的效应,因此该研究犯了“缺乏必要对照组”的错误,其结论的说理力不强。

**3.2 对照不全** 有些实验组有对应的对照组,而另一些实验组却缺乏对应的对照组。这种错误一般出现在涉及到考察交互作用的实验设计中,比如下面的例子。

例 3:《藻酸盐敷料与 mEGF 联合应用对难愈性创面 bFGF 影响的随机对照试验》<sup>[6]</sup>一文,主要是探讨联合应用藻酸盐敷料与冻干鼠表皮生长因子(mEGF)对难愈性创面碱性成纤维细胞生长因子(bFGF)表达的影响,并评价其疗效。该研究选择常规换药治疗 1 个月创面难愈合患者 18 例,随机分 3 组,每组 6 例:A 组藻酸盐敷料与 mEGF 联合治疗组;B 组 mEGF 治疗组(用普通纱布代替藻酸盐敷料);C 组常规治疗组(仅用凡士林纱布敷贴创面)。治疗 7、14、21、28 d 后评价创面愈合指数。结果见表 1。统计学分析用单因素方差分析及 *t* 检验。结论:藻酸盐敷料与 mEGF 联合应用治疗难愈性创面,能协同两者优势,较单用 mEGF 疗效更佳。

表 1 不同时相创面愈合指数变化( $\bar{x} \pm s, \%$ )

组别	7 d	14 d	21 d	28 d
A	37.83±4.22	85.00±9.10	94.33±6.28	99.17±1.00
B	32.17±4.49	73.17±7.44	85.17±9.13	95.50±5.54
C	23.83±5.27	54.83±8.45	64.83±6.11	85.00±7.80

辨析与释疑:该研究的主要目的是藻酸盐敷料与 mEGF 联合应用对难愈性创面的治疗效果的评价。从治疗方法上看,A、B、C 三组实际上涉及了 2 个实验因素,即“是否用 mEGF(2 水平)”和“纱布类型(3 水平)”(见表 2),应该有 6 个组,而实际实施时仅有 3 个组,无法考察这两个实验因素之间的交互作用是否具有统计学意义,因此,不能得出“藻酸盐敷料与 mEGF 联合应用治疗难愈性创面,能协同两者优势,较单用 mEGF 疗效更佳”的结论。该研究在实验设计上犯了“对照不全”的错误。

从上表可以看出,根据该实验研究的目的,要得出令人信服的结论,需要补充未做的 3 组实验。当补全表 2 中所有用“XX”表示的原始数据时,该实验设计叫做具有一个重复测量的三因素实验设计,“时间”是一个重复测量因素,“是否用 mEGF”和“纱布类型”是 2 个实验因素。在统计分析中应该采取“具有一个重复测量的三因素实验设计定量资料方差分析”来处理。

**3.3 对照过剩** 即设置多余的对照组。多余的对照不仅对于

表 2 不同时相创面愈合指数变化( $\bar{x} \pm s, \%$ )

纱布类型	是否用 mEGF	愈合指数			
		7 d	14 d	21 d	28 d
藻酸盐敷料	是(A)	37.83±4.22	85.00±9.10	94.33±6.28	99.17±1.00
普通纱布	是(B)	32.17±4.49	73.17±7.44	85.17±9.13	95.50±5.54
凡士林纱布	是	XX	XX	XX	XX
藻酸盐敷料	否(C)	XX	XX	XX	XX
普通纱布	否	XX	XX	XX	XX
凡士林纱布	否	XX	XX	XX	XX

验证研究假设没有任何作用,而且造成科研上的浪费,不利于进行正确的统计分析。

例 4:《斑蝥黄质对 D-半乳糖老龄大鼠骨质影响》<sup>[7]</sup>一文,为了研究斑蝥黄质对 D-半乳糖致衰老大鼠骨质的影响,进行了如下处理:将 45 只 6 周龄 Wistar 雄性大鼠随机分为 3 组(衰老模型组、斑蝥黄质组、青年对照组),每组 15 只,另设 16 月龄 Wistar 雄性大鼠 15 只作为老龄对照组。其中用 D-半乳糖对衰老模型组、斑蝥黄质组大鼠进行皮下注射用以建立衰老模型,建模成功后斑蝥黄质组大鼠用斑蝥黄质进行每日灌胃处理。青年对照组用以等量的生理盐水注射。5 个月后所有组一起进行骨常规参数测定并比较组间差异。设计和资料见表 3,请分析实验设计中分组是否合理?

辨析与释疑:原文设置了“老龄对照组”作为对照,犯了“对照过剩”的错误。根据原作者的研究目的,该研究涉及了两个实验因素:“建模与否”和“采用斑蝥黄质治疗与否”。需要回答的问题是:①建模组与未建模组差别有无统计学意义?②建模成功后斑蝥黄质治疗与否差别有无统计学意义?前者解决建模是否成功的问题,后者解决本文研究假设的问题。

值得注意的是,原作者所提及的“老龄大鼠”是由 D-半乳糖诱发的,而不是生理年龄达到了“老龄”。单设的老龄对照组涉及到“大鼠年龄(青年、老年)”这个因素,与本文研究目的没有直接关系。这属于重复设置对照组的错误,正确的做法是删除多余对照组即可。

如果未建模的青年对照组不进行斑蝥黄质治疗的话,该设计属于“多因素非平衡组合实验”(修改之后见表 4)。

表 4 不是一个标准的单因素 3 水平设计。对于这种定量资料,需要将 3 组拆分成 2 个不同组合,然后分别进行统计分析,即对照组与模型组为成组设计,模型组与斑蝥黄质组也为成组设计。

若设一个用斑蝥黄质治疗青年鼠组(表 5 中的第 3 行)在专业上允许的话,可将表 4 改为表 5 形式。

表 5 所对应的实验设计类型为一个标准的 2×2 析因设计,采用相应设计定量资料方差分析,可以明确回答“是否建模”与“是否用斑蝥黄质”之间的交互作用是否具有统计学意义

表 3 血清 MDA 含量和血液 SOD 活性( $\bar{x} \pm s, \text{rats}=15$ )

项目	青年对照组	衰老模型组	老龄对照组	斑蝥黄质组
MDA(nmol/ml)	6.72±0.52	14.52±0.49**	10.43±0.12**	7.22±0.20▲*
SOD(NU/mgpro)	121.32±5.21	94.42±4.16**	103.62±4.01**	110.55±4.11▲*

注:与青年对照组比较:\**P*<0.05, \*\**P*<0.01;与衰老模型组比较:▲*P*<0.01

义,能把专业问题解释得更加清楚。

表 4 各组血清 MDA 含量和血液 SOD 活性比较( $\bar{x}\pm s$ , rats=15)

组别	MDA(nmol/ml)	SOD(NU/mgpro)
对照组	6.72±0.52	121.32±5.21
模型组	14.52±0.49	94.42±4.16
斑蝥黄质组	7.22±0.20	111.55±4.11

表 5 各组血清 MDA 含量和血液 SOD 活性比较( $\bar{x}\pm s$ , rats=15)

是否用斑蝥黄质	是否建模	MDA(nmol/ml)	SOD(NU/mgpro)
否	否(对照组)	6.72±0.52	121.32±5.21
	是(模型组)	14.52±0.49	94.42±4.16
是	否	X	X
	是(斑蝥黄质组)	7.22±0.20	111.55±4.11

3.4 假对照 设立的对照组与实验设计目的不相符。

例 5:《Ilizarov 外固定架治疗痉挛型脑瘫屈膝畸形》<sup>[8]</sup>一文,对 21 例共 36 个畸形膝关节进行 Ilizarov 外固定架治疗,对术前和术后的疼痛指数、行动功能、屈膝改变程度等指标进行比较(如表 6),并且进行 Dimeglio 标准评分分析:36 个膝关节优 4 个,良 18 个,可 10 个,差 4 个。统计学处理:治疗前后的改变采用配对设计定量资料 *t* 检验进行统计处理。结论:正确使用 Ilizarov 技术矫治痉挛型脑瘫屈膝畸形,疗效满意,并发症少,是一种微创、安全、有效的治疗方法。

表 6 手术前后症状体征改变(分,  $\bar{x}\pm s$ )

随访项目	手术前	手术后	<i>t</i> 值	<i>P</i> 值
疼痛指数(级)	2.03±1.16	0.56±0.69	10.50	<0.01
行走功能指数(级)	2.25±0.87	0.28±0.45	15.44	<0.01
屈膝改变程度(度)	80.61±25.51	8.91±2.39	19.96	<0.01
关节活动范围(度)	66.50±26.97	117.78±11.51	14.79	<0.01

辨析与释疑:原文介绍了一种新方法在临床中的应用情况,并对治疗前后疗效进行了自身对照评价。该研究仅仅对某种疗法做了自身对照评价,没有把该疗法与其他疗法进行比较,也就是说,缺乏与同类病例中公认或常规的疗法进行对照研究,因此所得结论说服力不够强。该研究犯了“假对照”的错误,结合该研究的具体情况,至少有 3 个方面容易使人产生疑问:一是这种新疗法与其他方法疗效比较差别如何?二是这种新疗法所选受试对象与其他疗法的受试对象之间有无可比性?三是随着时间的推移,患者情况也会发生改变,怎样排除“自愈”对结果的干扰和影响?为了使结论具有说服力,这些问题必须得到有效地解决或合理解释。

3.5 盲目设立对照组 随意设置对照组,可能与研究目的无关。

例 6:《地震挤压伤对肢体骨折术后伤口愈合影响的临床观察》<sup>[9]</sup>一文,目的:观察地震挤压伤对术后伤口愈合影响。方

法:肢体在地震时曾被挤压而接受骨折内固定手术患者 28 例为观察组;以地震时肢体未被挤压接受骨折内固定手术的患者 27 例为对照 I 组;以同期本地骨折内固定手术患者 30 例为对照 II 组。分别检查尿常规、肌酸激酶,伤口并发症及伤口愈合情况。统计学处理:采用 SPSS 14.0,等级资料采用秩和检验,计数资料采用  $\chi^2$  检验。结果:各组尿常规、肌酸激酶无明显差异,而观察组伤口并发症、伤口愈合率与两个对照组均有差异。结论:地震挤压伤对术后伤口愈合有影响。资料见表 7。

表 7 3 组伤口愈合情况比较(例)

分组	甲级愈合	乙级愈合	丙级愈合
观察组	15	11	2
对照 I 组 <sup>▲</sup>	24	3	0
对照 II 组 <sup>*</sup>	29	1	0

注:与观察组比较,▲*P*<0.05,\**P*<0.05;两对照组比较,*P*>0.05

辨析与释疑:该研究设置了一组与研究目的不相干的对照组,犯了“盲目设立对照组”的错误。原文设置的对照 II 组所起的作用与对照 I 组大致相同,目的都是为了验证研究假设。而从均衡性原则而言,对照 I 组比对照 II 组更适合作为观察组的对照组。因此,该研究完全可以仅设立对照 I 组即可。

此外,为了达到研究目的,原作者仅仅以“挤压与否”作为实验的分组因素是不够的,比如挤压时间、挤压程度、是否合并感染等因素。该研究是一个比较复杂的问题,因此,如果要得到更科学的研究结论,需要综合考虑挤压相关的所有因素对疗效评价的影响。

参考文献

- [1] 胡良平. 口腔医学科研设计与统计分析. 北京:人民军医出版社,2007. 79-82.
- [2] 刘桂芬. 医学统计学. 第 2 版. 北京:中国协和医科大学出版社,2007. 167-169.
- [3] 胡良平. 统计学三型理论在实验设计中的应用. 北京:人民军医出版社,2006. 20-23.
- [4] 校佰平,李明,毛伟民,等. III 度开放性胫腓骨骨折外固定治疗策略. 中国骨伤,2008,21(4):289-290.
- [5] 王敦壮,李伟元,王晓波. 内外固定加中药治疗浮肘损伤. 中国骨伤,2008,21(5):358-359.
- [6] 毕擎,夏冰,朱丹杰,等. 藻酸盐敷料与 mEGF 联合应用对难愈性创面 bFGF 影响的随机对照试验. 中国骨伤,2007,20(10):659-662.
- [7] 裴凌鹏,惠伯棣,董福慧. 斑蝥黄质对 D-半乳糖老龄大鼠骨质影响. 中国骨伤,2008,21(8):613-616.
- [8] 胡炜,许世刚,曹旭,等. Ilizarov 外固定架治疗痉挛型脑瘫屈膝畸形. 中国骨伤,2008,21(12):922-923.
- [9] 陈福洪,陈泽,段恒琼,等. 地震挤压伤对肢体骨折术后伤口愈合影响的临床观察. 中国骨伤,2008,21(10):746-747.

(收稿日期:2009-04-27 本文编辑:李为农)