

· 继续教育园地 ·

实验设计不容忽视随机原则

胡良平, 陶丽新, 王琪, 毛玮

(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(6): 474-477 www.zggszz.com

一项科研工作或一篇学术论文, 衡量其科学性好坏的重要标志之一就是科研设计的质量。科研设计包括专业设计和统计研究设计两大方面。在统计研究设计(包括实验设计、临床试验设计和调查设计)中, 实验设计的理论和方法几乎是通用的。而实验设计的核心内容可以概括为: 三要素(受试对象、实验因素、实验效应)、四原则(随机、对照、重复、均衡)和实验设计类型(观察性研究设计和实验性研究设计)。统计学具有随机性特点。在自然界(特别是生物医学研究)中, 由于变异性的普遍存在, 又由于通常无法研究总体中的全部个体, 随机抽样研究是确保样本具有代表性的重要措施之一。要使样本很好地代表总体, 不仅要确保样本含量足够大, 更要强调用随机的方法从总体中抽取样本, 还要用随机的方法将样本分配到不同的处理组中去, 否则, 研究结论的可信度将会大大降低。本文针对《中国骨伤》杂志 2008 年第 1 期至第 12 期中出现的违反四原则的典型案例分析, 以便对临床科研工作者有所帮助。

1 随机原则的概念与作用

随机原则的概念: 在选取样本时, 应确保总体中任何一个个体都有同等的机会被抽取进入样本; 在分配样本时, 应确保样本中任何一个个体都有同等的机会被分入任何一个组中去。

随机原则的作用: 其作用就是使样本具有极好的代表性, 使各组受试对象在重要的非实验因素方面具有极好的均衡性, 提高组间实验资料的可比性。

随机化应该体现在以下几个方面:

(1) 抽样随机: 每一个符合条件的实验对象参加实验的机会相同, 即总体中每个个体有相同的会被抽到进入样本之中。

(2) 分组随机: 每个实验对象分配到不同组(通常为对照组、不同处理组)的机会相同。

其中, 第(1)条: 保证所得到的样本具有代表性, 使实验结论具有普遍意义; 第(2)条: 保证各组间实验对象尽可能均衡一致, 以提高各组间的可比性^[1-2]。

2 实现随机化的一般方法

实现随机化的方法有多种, 可以采用“抽签”法、查“随机数字表”、“随机排列表”、查“用计算机产生的伪随机数字表”或直接由计算机程序来实现^[3]。

2.1 完全随机化的一般步骤如下

第一步, 将受试对象编号, 将编号按顺序写成一排。

第二步, 事先规定分组的规则。如分 2 组时, 可规定遇到随机数字为偶数时将对应的受试对象分入实验组, 遇到随机数字为奇数时将对应的受试对象分入对照组(反过来规定也可以); 再如分 3 组时, 可事先规定, 凡随机数字除以 3 得余数为 0 者分入第 1 组, 余数为 1 者分入第 2 组, 余数为 2 者分入第 3 组。当然, 也可以规定其他的分组规则, 但规则必须事先确定下来, 一旦确定不应随意改动。

第三步, 从上述 3 种随机数字表中任意指定的位置开始向后(或向前)抄录随机数字, 依次写在各编号之下。注意: 舍弃不符合要求的随机数字(如随机数字超过了编号所对应的数字)。

第四步, 根据抄录的随机数字按事先确定的分组规则分组。当各组样本含量不等时最好再用随机的方法进行调整, 尽可能使各组的样本含量相等或接近相等(从统计计算角度看, 各组样本含量相等时误差较小)。若采用的是“随机排列表”, 一般来说, 分组的结果会使各组样本含量相等, 这样麻烦事就少多了。当然, 现在可以运用计算机程序直接实现随机化分组, 使随机化的过程大为简化, 非常方便。

2.2 介绍临床试验中一种简便易行的随机化分组方法 在临床试验研究中, 常习惯将患者分为试验组与对照组, 患者来医院就诊, 若将先来就诊的分入试验组, 后来就诊的分入对照组, 是不正确的。因为患者来医院就诊, 在患病严重程度、患病时间等重要非实验因素方面不一定是均衡的, 可能带有某种程度上的“聚集性”, 即某一段时间上来的患者多数为重病患者, 另一段时间上来的多数为轻病患者; 若规定第 1 个来就诊者分入试验组, 第 2 个来就诊者分入对照组, 这样交叉进行分组, 最后也可能两组患者在某些重要非实验因素方面相差悬殊, 组间缺乏均衡性。

有一种简便易行的随机化分组方法叫“按不平衡指数最小的分配原则”^[4]分组。这个分组方法的做法如下: 根据专业知识选取几个拟加以控制的重要非实验因素, 假定 1 个是患者的“性别(分为男、女)”, 另一个是患者的“病情(分为轻、中、重)”。将先来的两位患者在试验组与对照组各放 1 人, 记下他们的性别和病情, 记分方法是每个因素的每个水平出现 1 次记 1 分, 计算两组各因素对应水平的得分之差的绝对值, 最后求出绝对值之总和, 称此“和”为两组患者在两个重要非实验因素上的不平衡指数。若再来第 3 位患者, 分别依次将此患者放入试验组、对照组各 1 次, 每次都根据他(或她)的性别、病情累加到原有患者的基础之上, 可以得到两个不平衡指数, 取

不平衡指数最小的那种分组方法, 这样第 3 位患者的分组就定下来了, 用同样的方法去分配以后来的该病患者, 直到两组有了事先规定的样本含量时就停止。现举例说明如下:

假定在对肩周炎患者分 2 组时考虑病情(轻、中、重)、患病时间(短、长)和日常运动量(少、多)3 个重要的非处理因素, 再假定治疗组和对照组各有了 1 位患者, 他们的基本情况如表 1 所示。又来了一位新患者, 病情重、患病时间短、日常运动量多, 问将此患者分入哪个组中去好?

表 1 2 例患者的基本情况及分组后的平衡程度

实验	例数							
	病情: 轻	中	重	患病时间: 短	长	运动量: 少	多	合计
治疗组	0	1	0	0	1	1	0	
对照组	1	0	0	1	0	1	0	
差的绝对值	1	1	0	1	1	0	0	4

注: 表中合计值“4”就是“不平衡指数”, 下同

试着将新患者分别分入治疗组和对照组, 取“不平衡指数”较小者所对应的分配方案, 分别参见表 2 和表 3。

表 2 将新患者分入治疗组后的平衡程度的探索

实验	例数							
	病情: 轻	中	重	患病时间: 短	长	运动量: 少	多	合计
治疗组	0	1	1	1	1	1	1	
对照组	1	0	0	1	0	1	0	
差的绝对值	1	1	1	0	1	0	1	5

表 3 将新患者分入对照组后的平衡程度的探索

实验	例数							
	病情: 轻	中	重	患病时间: 短	长	运动量: 少	多	合计
治疗组	0	1	0	0	1	1	0	
对照组	1	0	1	2	0	1	1	
差的绝对值	1	1	1	2	1	0	1	7

因 $5 < 7$, 故本例应将新患者分入治疗组。若再有新患者, 分配方法相同。假如有 $k (k \geq 2)$ 个治疗组, 可分别与对照组比较, 分别求 $(k-1)$ 个表中的“不平衡指数”, 取其中最小者对应的分配方案。

3 错误案例分析与释疑

3.1 选择受试对象时没有遵循随机原则

3.1.1 回顾性研究在选择受试对象时通常无法遵循随机原则

例 1, 2000 年至 2006 年收治 63 例胫骨平台骨折, 现对内固定手术治疗进行回顾性分析, 提出各种具体治疗方法, 以期进一步提高手术的远期效果。本组 63 例, 男 37 例, 女 26 例; 年龄 19~54 岁^[5]。

例 2, 回顾性分析三通道髓芯钻孔减压技术加自体骨髓干细胞、DBM 植入治疗早期股骨头坏死疗效, 并探讨其适应证。根据股骨头坏死史夕法尼亚大学分型方法对 87 例 103 髌进行三通道髓芯钻孔减压技术加自体骨髓干细胞、DBM 植入治疗早期股骨头坏死, 其中男 71 例(86 髌)女 16 例(17 髌); 平均年龄 47 岁; 平均病程 1.3 年^[6]。

例 3, 回顾性分析一组多阶段椎弓根螺钉内固定治疗中上胸椎骨折脱位的临床资料, 探讨多节段椎弓根螺钉内固定的治疗效果及理论依据。胸椎骨折脱位 23 例, 男 21 例, 女 2 例; 年龄 20~47 岁, 平均 33.5 岁^[7]。

例 4, 对 2005 年至 2007 年的 32 例糖尿病足截肢(趾)患者临床资料做了回顾性分析, 并对其进行简单讨论。本组 32 例, 男 25 例, 女 7 例; 年龄 51~78 岁, 平均 65.1 岁, 60 岁以上 23 例; 右足 22 例, 左足 10 例。32 例均为单侧截肢(趾), 糖尿病病史不足 1 年者 3 例, 1~5 年 11 例, 5 年以上 18 例^[8]。

【对差错的辨析与释疑】上述 4 个案例的原文作者选择受试对象时, 均选择了既往治疗过的患者, 这种类型的研究属于“回顾性”研究, 在受试对象的选择上无法满足随机的要求, 正确方法应该是从患病的总体中随机抽取需要数量的患者, 而回顾性研究限定了时间段, 有时还会限定某家医院或者某个地区, 无法遵循随机原则。这些患者在年龄、病程、疾病的严重程度、损伤部位等许多方面存在较大的差异, 而“回顾性”研究是无法解决这些问题的。因此这种“回顾性”研究的结论的可信度很低, 有些甚至是错误的!

目前临床科研存在的误区就包括这一种情况, 而且这种情况是相当严重的。某些临床医师认为临床科研很简单, 就是总结一下日常的临床工作经验, 导致很多临床科研课题的科研设计质量不高, 结论没有说服力。临床科研也应该考虑多做一些前瞻性研究, 因为随机对照临床实验(RCT)最具有说服力。在不违反伦理道德的前提下, 应尽可能多地开展大规模前瞻性临床试验研究, 严格按照试验设计的 4 个基本原则(即随机、对照、重复和均衡)去实施试验, 选择特异性强、灵敏度高的观测指标, 加强临床试验过程中的质量控制, 正确运用统计分析方法处理实验资料, 方可得出正确的统计和专业结论^[9]。

3.1.2 在涉及受试对象的选择时没有提及随机原则

例 5, 收集 32 例前交叉韧带(ACL)损伤病例, 单纯 ACL 损伤 29 例, 与后交叉韧带(PCL)同时损伤 3 例。分为 2 组, 同种异体肌腱重建 15 例, 男 12 例, 女 3 例, 年龄 23~61 岁; 自体腓肠肌腱重建 17 例, 男 11 例, 女 6 例, 年龄 17~57 岁^[10]。

例 6, 收集 20 具新鲜成人上颈椎标本, 分 3 组, 模拟齿突 II 型骨折模型, 骨折线为前下后上, 截骨角度分别为 0° 6 具, 17° 8 具, 25° 6 具, 并进行齿突螺钉固定。进行旋入力矩、最大轴向拔出力和刚度测试^[11]。

例 7, 2003 年 12 月至 2006 年 12 月采用自行设计获国家专利的耻骨联合固定板治疗耻骨联合分离 18 例, 取得满意疗效。本组 18 例, 男 12 例, 女 6 例; 年龄 22~64 岁, 平均 38 岁。车祸伤 10 例, 挤压伤 4 例, 高处坠落伤 2 例, 其他伤 2 例。合并耻、坐骨骨折 4 例, 骶髂关节脱位 2 例, 骶骨骨折 1 例, 髌臼骨折 2 例^[12]。

例 8, 改进夹板固定治疗桡骨远端粉碎性骨折, 观察其疗效并与传统夹板固定进行放射学比较。92 例桡骨远端粉碎骨折患者, 男 38 例, 女 54 例; 年龄 23~82 岁, 平均 63 岁。按 AO 分型: A3 型 15 例, C1 型 40 例, C2 型 26 例, C3 型 11 例^[13]。

例 9, 观察研究三维复位固定器治疗胫腓骨骨折的临床疗效。胫腓骨骨折患者 21 例, 男 12 例, 女 9 例; 年龄 18~72 岁, 平均 46 岁。开放性骨折 5 例, 闭合骨折 16 例。上段骨折 4 例,

中段骨折 8 例,下段骨折 9 例。斜形骨折 10 例,粉碎骨折 8 例,多段骨折 3 例^[14]。

【对差错的辨析与释疑】上述 5 个案例的原文作者在选择受试对象时,均没有提及随机原则,其在选择受试对象的过程中是否遵循了随机原则,我们很难知道。换句话说,原作者可能在选择受试对象时忽视了随机原则,从而导致试验研究的结论不令人信服。

3.1.3 随意选择受试对象,违反随机原则

例 10,自 2006 年 2 月至 2007 年 2 月,采用射频热凝技术共治疗腰椎间盘突出症患者 86 例,现选择资料齐全的腰椎间盘突出症 28 例总结如下^[15]。

例 11,探讨载脂蛋白 A1、B 基因多态性对非创伤性股骨头坏死(ANFH)发生的影响。应用聚合酶链反应对中国北方汉族 143 例 ANFH 患者和 92 例正常人分别扩增含 ApoA1 基因启动子-75 bp 和第一内含子+83 bp 及 ApoB 基因 EcoRI、Xba I 和 3'-VNTR 的 DNA 片段,限制性内切酶酶切扩增产物,琼脂糖凝胶电泳分离基因多态性^[16]。

【对差错的辨析与释疑】以上 2 个案例的作者在选择受试对象时没有遵循随机原则。例 10 中,选择的患者是从 86 例中选择了资料齐全的 28 例,并不是随机选择患者,而资料齐全与否可能是影响结果或作出某些评价的一个重要非实验因素,而原作者人为地把这一因素去掉了,没有考察它对结果的影响,使得出的结论说服力不强。例 11 中,作者在选择受试对象时也没有遵循随机原则,并不是从受试对象的总体中抽取样本,而是限定了试验组是中国北方汉族的 143 例患者,抽取样本时并没有遵循随机原则。

3.2 对受试对象进行分组时没有遵循随机原则

3.2.1 在对受试对象分组时,用“随意”取代“随机”

例 12,探讨电针对骨关节炎的治疗作用。将 40 例膝骨关节炎患者按就诊顺序随机分为 2 组,每组 20 例^[17]。

例 13,2005 年 3 月至 2006 年 8 月骨科门诊、急诊收治的 24 h 内新鲜各类桡骨远端骨折 90 例,男 28 例,女 62 例;年龄 61~91 岁,平均 63 岁。将 90 例患者按就诊序号分为中药组、西药组、空白对照组^[18]。

例 14,2001 年 1 月至 2005 年 12 月在关节镜引导下应用 GSH 治疗股骨远端骨折 49 例,与同期切开复位 GSH 钉固定 32 例比较,取得满意疗效。81 例股骨远端骨折患者根据入院顺序分组,分别采用切开复位 GSH 钉固定和关节镜引导下 GSH 钉治疗^[19]。

例 15,2004 年 8 月至 2006 年 6 月收治的腰椎间盘突出症患者 82 例。2004 年 8 月至 2005 年 5 月采取单纯 PLDD 治疗 36 例(激光组),2005 年 6 月至 2006 年 6 月采用 PLDD 术后组合腰椎牵引治疗 46 例(组合组)^[20]。

【对差错的辨析与释疑】以上 4 个案例中,分组时是按照就诊顺序分组,虽然例 12、13 和 14 是按照就诊顺序随机分组,也不能保证分组的效果好,而例 15 则是以 2006 年 6 月为界,将患者分为 2 组。这两种分组方法都是不科学的,因为患者来医院就诊在患病严重程度、患病时间等重要的非实验因素方面不一定是均衡的,可能带有某种程度上的“聚集性”,即某一段时间上来的患者多数为重病患者,另一段时间上来的

多数为轻病患者;若规定第 1 个来就诊者分入试验组,第 2 个来就诊者分入对照组,这样交叉进行分组,最后也可能两组患者在某些重要非实验因素方面相差悬殊,组间缺乏均衡性。正确的方法是用上面介绍的“按不平衡指数最小的分配原则”进行分组。

3.2.2 用完全随机取代分层随机

例 16,探讨钳持端提回旋手法复位经皮逆行穿针内固定治疗锁骨骨折的疗效及安全性,按随机数字表法分组。治疗组男 75 例,女 26 例;年龄 18~65 岁,平均(40.5±2.4)岁。骨折类型:粉碎性 84 例,长斜形 9 例,短斜形或横断形 8 例。骨折移位程度:骨折端重叠 9~21 mm,平均(12.0±0.7) mm。对照组男 69 例,女 31 例;年龄 19~65 岁,平均(39.1±3.6)岁。骨折类型:粉碎性 88 例,长斜形 6 例,短斜形或横断形 6 例。骨折移位程度:骨折端重叠 8~21 mm,平均(12.0±0.6) mm。两组经 χ^2 检验或 *t* 检验,在性别构成、年龄分布、分型组成、骨折移位程度等方面无统计学差异 ($P>0.05$)^[21]。

例 17,收集 2001 年 9 月至 2006 年 7 月 176 例颈椎病手术患者,均行前路减压植骨钛板内固定术,男 98 例,女 78 例;年龄 36~78 岁,平均 50.2 岁。按随机分组原则,于手术前 1 天采用掷钱币法对患者进行随机分组。A 组:缝合椎前筋膜 89 例,男 51 例,女 38 例;年龄 36~78 岁,平均 48.2 岁。B 组:不缝合椎前筋膜 87 例,男 47 例,女 40 例;年龄 42~73 岁,平均 47.8 岁。两组患者性别、年龄经统计学处理差异无统计学意义,有可比性^[22]。

【对差错的辨析与释疑】在例 16 中,作者将患者按照随机数字表法分组,即按照完全随机法进行了分组;在例 17 中,原文作者采用掷钱币法对患者进行随机分组。虽然原作者进行检验之后,认为在性别构成、年龄分布、分型组成和骨折移位程度等方面无统计学差异,但在样本并非足够大时,由常识可知,完全随机化分组能达到如此理想的均衡效果的概率是比较小的,因此,该两例的分组结果令人怀疑。比较科学的做法是采用“分层随机化”,即根据前述的几个重要非实验因素将全部符合纳入标准不符合排除标准的患者分成很多小组,每个小组中的患者性别、年龄、型别和骨折端重叠情况相同或十分接近,再将每个小组中的患者完全随机地均分入试验组和对照组,以使各个重要的非实验因素在组间比较均衡,得到的结论才令人信服。

综上所述,在医学科研设计中,实际工作者对统计学中随机原则的遵循确实还存在很多问题,有待于进一步加强学习和改进。好的开始是成功的一半,如果实验设计正确,即使统计分析方法选错了,还可以予以纠正。然而,如果实验设计就发生了错误,则意味着整个实验的失败。不言而喻,随机原则是科学实验设计中的重要环节,是绝对不容忽视的!

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. 78.
[2] 胡良平. 医学统计实战练习. 北京:军事医学科学出版社,2007. 2-3.
[3] 方积乾. 医学统计学与电脑实验. 第 2 版. 上海:上海科学技术出版社,2001. 7.

[4] 胡良平. 检验医学科设计与统计分析. 北京: 人民军医出版社, 2004. 5.

[5] 黄海涛, 周中, 江宁. 内固定手术治疗胫骨平台骨折临床分析. 中国骨伤, 2008, 21(9): 696-697.

[6] 季卫锋, 丁伟航, 马镇川, 等. 三通道髓芯钻孔减压加 DBM、自体骨髓干细胞治疗早期股骨头坏死. 中国骨伤, 2008, 21(10): 776-778.

[7] 徐丽明, 顾锐, 林野, 等. 多节段椎弓根螺钉内固定治疗中上胸椎骨折脱位. 中国骨伤, 2008, 21(8): 603-605.

[8] 赵波, 贺西京, 李晟. 糖尿病足截肢(趾)32 例临床分析. 中国骨伤, 2008, 21(7): 546-547.

[9] 胡良平. 科研课题的研究设计与统计分析——错误案例辨析与释疑(第 1 集). 北京: 军事医学科学出版社, 2008. 3.

[10] 刘红光, 陈述祥, 赵崇达, 等. 膝关节镜下自体腘绳肌腱和同种异体肌腱移植重建前交叉韧带对比观察. 中国骨伤, 2008, 21(4): 267-269.

[11] 方弘伟, 李建军, 吴增晖, 等. 前下后上型齿突骨折螺钉内固定的生物力学研究. 中国骨伤, 2008, 21(4): 279-281.

[12] 黄儒收, 贾怀海, 杨光, 等. 耻骨联合固定板治疗耻骨联合分离. 中国骨伤, 2008, 21(4): 311.

[13] 陈民, 林学波, 王寰, 等. 改进夹板固定治疗桡骨远端粉碎性骨折的临床疗效对比研究. 中国骨伤, 2008, 21(2): 87-89.

[14] 阎兆东, 赵勇, 张泰标, 等. 三维复位固定器治疗胫腓骨骨折. 中国骨伤, 2008, 21(2): 97-98.

[15] 杨利学, 谭龙旺, 李小群, 等. CT 引导下射频术对腰椎间盘突出出症髓核即时回缩效应的临床观察. 中国骨伤, 2008, 21(4): 319.

[16] 汪晓燕, 牛晓红, 陈卫衡, 等. 载脂蛋白 A1、B 基因多态性对非创伤性股骨头坏死发生的影响. 中国骨伤, 2008, 21(2): 99-102.

[17] 吴志宏, 包飞. 电针治疗膝骨关节炎临床对照试验. 中国骨伤, 2008, 21(3): 170-172.

[18] 陈述祥, 康乐, 陈辉强. 夹板外固定结合中药治疗老年人桡骨远端骨折的临床对照试验. 中国骨伤, 2008, 21(3): 181-183.

[19] 李杰. 关节镜下与切开逆行交锁髓内钉治疗股骨远端骨折临床疗效比较. 中国骨伤, 2008, 21(2): 145-146.

[20] 沈进稳, 瞿杭波, 童培建. 牵引疗法在激光腰椎间盘汽化减压术后的应用价值. 中国骨伤, 2008, 21(8): 632-633.

[21] 毕宏政, 杨茂清, 谭远超. 钳持端回旋手法复位经皮逆行穿针内固定治疗锁骨骨折的随机对照试验. 中国骨伤, 2008, 21(7): 490-493.

[22] 董胜利, 陈海啸. 缝合椎前筋膜预防颈前路术后吞咽困难的临床研究. 中国骨伤, 2008, 21(8): 606-607

(收稿日期: 2009-03-12 本文编辑: 李为农)

《中国骨伤》编辑委员会名单

名誉主编:(按首字汉语拼音字母顺序为序)

陈可冀(中国科学院院士) 葛宝丰(中国工程院院士) 沈自尹(中国科学院院士)
王澍寰(中国工程院院士) 吴咸中(中国工程院院士) 钟世镇(中国工程院院士)

顾问:(按首字汉语拼音字母顺序为序)

陈渭良 丁继华 冯天有 顾云伍 胡兴山 蒋位庄 孔繁锦 黎君若 李同生 梁克玉
刘柏龄 孟和 施祀 时光达 石印玉 孙材江 袁浩 赵易 朱惠芳 朱云龙
诸方受

主编:董福慧

副主编:(按首字汉语拼音字母顺序为序)

敖英芳 白人骁 杜宁 金鸿宾 李为农(常务) 吕厚山 邱勇 孙树椿 王岩
王满宜 卫小春

编委委员:(按首字汉语拼音字母顺序为序)

敖英芳 白人骁 毕大卫 陈仲强 董健 董福慧 董清平 杜宁 樊粤光 葛尊信
郭万首 郭卫 何伟 胡良平 胡兴山 金鸿宾 雷仲民 李德达 李盛华 李为农
李无阴 刘金文 刘兴炎 刘忠军 刘仲前 罗从风 吕厚山 马真胜 邱勇 阮狄克
沈霖 沈冯君 石关桐 孙常太 孙树椿 孙天胜 谭明生 谭远超 王岩 王爱民
王和鸣 王坤正 王满宜 王序全 王拥军 韦贵康 卫小春 肖鲁伟 徐荣明 杨小平
姚共和 姚树源 余庆阳 袁文 詹红生 张俐 张保中 张春才 张功林 张连仁
张英泽 赵平 赵建宁 赵文海 郑忠东 钟广玲 周卫 朱立国 朱振安 邹季
顾华(美国) John W. McDonald(美国)