

• 继续教育园地 •

# 重复原则在骨科实验设计中的应用

胡良平, 王琪, 陶丽新, 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(10): 800-802 www.zggszz.com

实验设计是科学研究的一个重要环节, 良好的实验设计是顺利进行科学研究和数据分析的先决条件, 也是获得预期结果的重要保证<sup>[1]</sup>。科学研究的目的, 是为了揭示事物的客观规律。研究的价值首先决定于研究的目的, 而研究的结果是通过一定方法步骤而获得的。一个周密的实验研究全过程, 通常要联系质的特性, 研究量的表现, 所以从统计方面讲, 主要应考虑随机、对照、重复、均衡等问题, 这是实验研究的一般原则。实验要求一定重复数(样本例数), 其目的是使均数逼真, 并稳定标准差, 只有这样来自样本的统计量才能代表总体的参数, 统计推断才具有可靠的前提<sup>[2]</sup>。

## 1 重复的含义

重复通常有三层含义, 即“重复取样”、“重复测量”和“重复实验”<sup>[3]</sup>。从同一个样品中多次取样, 测量某定量指标的数值, 称为“重复取样”, 目的是看各标本中某定量观测指标含量的分布是否均匀; 对接受某种处理的个体, 随着时间的推移, 对其进行多次观测, 称为“重复测量”, 目的是看定量指标随时间推移(或部位改变)的动态变化情况; 实验设计中所讲的重复原则指的是“重复实验原则”, 即在相同的实验条件下的独立重复实验的次数足够多。这里的“独立”是指要用不同的个体或样品做实验, 而不是在同一个体或样品上做多次实验。

## 2 重复原则的作用

当人们关心的实验结果是一个不能事先准确判定的某种结局时, 一般称其为“随机变量”的一种具体表现, 如某种药物治疗某种骨病的患者, 其结果可能是“治愈、显效、好转、无效、死亡”之一, 事先谁也说不准。具有随机性结果的实验研究统称为“随机实验”, 要想比较正确地反映随机实验结果出现的一般规律, 必须进行大量的独立重复实验。又如, 投掷 1 枚质地均匀的硬币实验, 看其结果是正面朝上还是反面朝上, 投掷 1 次或 10 次, 其结果都是随机的, 而投掷 10 万次, 则正面朝上或反面朝上的机会接近相等, 约为 0.5。当然, 在实际的科研工作中, 在一个特定的条件下不可能做无数次独立重复实验, 做好结合具体情况作出合理的估计, 一般来说, 不做重复实验或仅重复 2、3 次实验, 其可靠性是值得怀疑的<sup>[4]</sup>。

## 3 科研课题中样本含量严重不足的实例

例 1 《骨髓间质干细胞体外定向诱导分化为软骨细胞的实验研究》<sup>[5]</sup>一文, 抽取兔髌骨骨髓液(3 月龄新西兰兔, 雌雄不限, 体质量 2~2.5 kg), 经梯度离心法和贴壁法进行体外培养, 贴壁细胞传代, 取第 3 代细胞在培养基中添加软骨分化诱导剂[转化生长因子(TGF- $\beta_2$ ) 10 ng/ml、地塞米松 10<sup>-7</sup> mol/L、

维生素 C 50  $\mu$ mol/L], 经 7、14、21 d 诱导培养后, 倒置显微镜观察细胞形态, 免疫组织化学染色检测软骨特异性 II 型胶原表达。将软骨细胞与软骨支架材料——聚磷酸钙纤维/左旋聚乳酸(CPP/PLLA)聚合, 1 周后终止培养, 扫描电镜观察细胞黏附情况。诱导后细胞体外扩增能力显著降低, 细胞形态由成纤维样梭形向多角形、多边形或类圆形转变, 诱导 21 d 后细胞形态变化最为显著, II 型胶原免疫组化染色深而均匀。诱导后的骨髓间充质干细胞可在支架材料内良好黏附生长。结论为体外培养的骨髓间充质干细胞可定向诱导分化为软骨细胞, 分泌软骨细胞特异性基质, 可用作软骨组织工程的种子细胞。

例 2 《间歇性压力培养环境对兔骨髓基质干细胞增殖的影响》<sup>[6]</sup>一文, 以 1 只出生 1 个月的新西兰兔的骨髓基质干细胞为实验材料, 将生长良好的第 3~4 代 MSC 按 1 $\times$ 10<sup>4</sup> 个/孔随机接种于 4 块 24 孔培养板, 14 h 后细胞完全贴壁, 换新鲜含 10% 胎牛血清的 DMEM(含 1 U/L 青霉素及 100 ms/L 链霉素)。按压力施加类型不同分为持续性和间歇性压力, 每块培养板的细胞为 1 个大组, 随机将 4 块培养板分为 4 个大组: 对照组、20 kPa 间歇组、60 kPa 间歇组、100 kPa 间歇组。每大组按观察期分为 4 个小组即: 1、3、5、7 d 组, 每小组样本(细胞)为 6 孔。对照组不予加压, 20、60、100 kPa 间歇组分别在每天施以 20、60、100 kPa 的压力 5 min 2 次。结果: MSC 在 60、100 kPa 间歇性压力培养环境下 MTT 反应的 OD 值小于对照组; 压力值越大, OD 值越小(P<0.05)。而 20 kPa 间歇性压力情况可显著促进 MSC 增殖, 并随时间增加而增强。在实验第 7 天时, 实验组 OD 值高于对照组(P<0.05), 差异有显著性意义。结论为关节置换术后早期下地活动产生的较大应力, 会明显抑制骨髓腔内 MSC 的增殖, 不利于骨的重建愈合, 临床应当避免。然而, 较小的间歇性压力对 MSC 增殖有明显的促进作用, 提示关节置换患者可在较小压力范围(<20 kPa)内进行早期关节被动活动训练, 有利于关节术后骨的重建愈合。

辨析: 上述 2 例的实验设计不正确。研究者在整个研究过程中所用到的髌骨骨髓液或骨髓基质干细胞只来自一只新西兰兔, 也就是说, 只用了 1 只兔的髌骨骨髓液或骨髓基质干细胞来做实验, 即样本含量 n=1。由于生物体个体差异的客观存在, 样本含量 n=1 的样本无论如何都是不具有代表性的, 不能代表任何一个总体。因此, 本研究的实验设计严重违反了重复原则。实际上, 研究者在本研究中是误用“重复取样”来代替“独立重复实验”。

例 3 《冲击载荷作用下骨盆三维有限元分析及其生物力

学意义》一文<sup>[7]</sup>,选取 1 名女性志愿者(39 岁)进行实验,采用计算机仿真模拟方法,将所构建骨盆三维实体模型导入三维有限元分析软件 ANSYS 7.0,计算单侧髌前上棘和单侧髌骨正后方部位冲击载荷作用下骨盆的力学行为表现,冲击载荷为 0~8 000 N,峰值 8 000 N,时间自 0~40 ms,分析主应力值、应力分布情况以及主应力方向上骨盆单元的位移。冲击载荷作用于单侧髌前上棘时,0~10 ms 主要的应力沿髌后上棘到髌前上棘分布,在冲击后期 10 ms 以后,髌髌关节、髌白和耻骨支都会产生较大的应力分布;作用于单侧髌骨正后方部位时,0~20 ms 主要的应力沿髌骨纵行传导,应力分布并没有在冲击峰值 10 ms 时达到最大,而是从 20 ms 之后,应力开始向髌髌关节、双侧耻骨支、坐骨支以及髌白等部位传导并可见明显的应力分布。

辨析:生物医学研究的现象常带有变异性,只有在相同实验条件下进行多次独立重复实验,随机现象的变化规律性才能正确地显露出来。在该研究中,只选取了 1 名 39 岁的女性志愿者,分析出来的结果只是在该名志愿者本身机体上体现得个性,并不能体现共性,其相关结论不能随意外推。

例 4《强骨宝方对外培养成骨细胞影响的实验研究》<sup>[8]</sup>一文,采用胰蛋白酶——II 型胶原酶消化法从 2 只 1~2 日龄 SD 乳鼠颅盖骨中分离出成骨细胞,鉴定细胞并传代培养后,分为对照组与 4 个实验组,实验组分别用终浓度 100、50、10、5 μg/ml 的强骨宝方提取液加入成骨细胞培养体系,对照组用不含强骨宝方提取液的培养基培养,应用 MTT 比色法、ALP 含量测定、矿化结节形成等分别观察其对成骨细胞增殖、分化及矿化能力的影响。所得计量资料均采用  $\bar{x} \pm s$  表示,应用 SPSS 11.0 软件进行统计分析,多组之间比较先采用 *F* 检验,而后进行 *q* 检验,以 *P*<0.05 为差异有显著性意义。100、50、10 μg/ml 浓度的强骨宝方提取液均具有促进体外成骨细胞增殖、分化与矿化的作用,以 100 μg/ml 与 50 μg/ml 促进作用最强。结论为每毫升含生药 2 g 的强骨宝方提取液, pH=7.0, 50 μg/ml 的添加浓度可能最适合于体外培养成骨细胞的增殖、分化及矿化成骨能力。见表 1-2。

表 1 不同浓度的强骨宝方提取液对成骨细胞增殖功能影响的比较( $\bar{x} \pm s, n=6$ )

组别	成骨细胞分泌 ALP 值	
	时间(h):	
	48	72
100 μg/ml 组	0.367 7±0.012 9	0.576 3±0.032 7
50 μg/ml 组	0.399 3±0.031 5	0.625 0±0.020 1
10 μg/ml 组	0.371 3±0.018 3	0.562 0±0.043 6
5 μg/ml 组	0.349 5±0.020 4	0.528 4±0.029 3
对照组	0.355 0±0.019 8	0.513 7±0.017 6

辨析:原作者只用了 2 只乳鼠,其所培养传代的细胞均来自这 2 只乳鼠,实验分 5 组,每组 *n*=6,这种做法违反了实验设计的重复原则,真正的样本含量是乳鼠数量 2,5 组的样本均来自于 2 只乳鼠的细胞,这样的实验代表性不够好,不能全面真实反映出这种动物间的个体差异对实验结果的影响,应该增加样本数量,分几批重复此实验,以便充分揭示药物剂量对观测结果的影响。

表 2 不同浓度的强骨宝方提取液对成骨细胞分泌 ALP 与矿化功能影响的比较( $\bar{x} \pm s, n=6$ )

组别	OD 值(king unit/100 ml)	矿化结节数(n/hole)
100 μg/ml 组	2.789 2±0.168 9	12.67±2.61
50 μg/ml 组	2.802 0±0.181 9	12.50±2.74
10 μg/ml 组	2.692 5±0.221 6	10.17±2.56
5 μg/ml 组	2.351 6±0.192 6	7.83±2.48
对照组	2.308 5±0.125 6	6.67±2.73

例 5《三角纤维软骨复合体解剖及生物力学研究》<sup>[9]</sup>一文,取 2 个肱骨中段以上的上肢标本,去除所有的前臂伸屈肌腱,前臂旋转中立位下固定桡骨,对尺骨小头施加 20 N 的掌背侧拉力,测量尺骨相对于桡骨的位移。然后先后切断背侧和掌侧桡尺韧带,相同的作用力下测量尺骨相对于桡骨的位移。对 DRUJ 稳定性的生物力学作了初步的研究表明,在 20 N 拉力作用下,切断背侧桡尺韧带,尺骨相对桡骨的背侧位移明显增加;切断掌侧桡尺韧带,保持完整的背侧桡尺韧带,尺骨相对桡骨的掌背侧位移都明显增加;而掌背侧桡尺韧带都切断导致 DRUJ 明显不稳(图 1)。

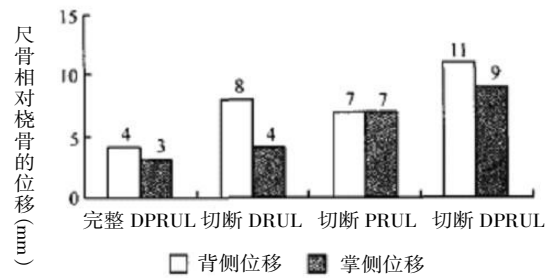


图 1 尺骨相对桡骨的位移

辨析:在大多数情况下,我们要在相同实验条件下做足够多次的独立重复实验,来找出比较可靠的客观规律。原文实验只取了 2 个样本,实验涉及“是否切断 DRUL”、“是否切断 PRUL”2 个实验因素,只利用 2 个样本完成多因素研究,严重违反了实验设计的重复原则,使得实验的结论可信度不高。

重复是指各处理组(对照在实验研究中也常被看作是一种处理)的实验单位都要有一定的数量,即要考虑样本含量的大小。样本太小,使应有的差别显示不出来,难以获得正确的研究结果,结论亦缺乏充分的依据;但样本太大,也会增加实际工作中的困难,造成人力、物力、时间的不必要浪费。因此,从事一项实验设计之前,很有必要考虑到样本含量问题<sup>[10]</sup>。

#### 4 什么是样本含量

实验设计所包括的各组内重复实验次数之和,称为样本大小或样本含量。在一项具体的调查或实验研究中,究竟应该用多大的样本含量合适呢?一般来说,若从严格定义的总体中随机抽样,样本含量越大越好。但事实上,这样是行不通的。因为样本越大,所需花在调查或实验上的精力、人力和物力都会很大,有时需要的时间很长。一旦人力和物力等不足,工作开展起来必然会粗枝大叶,数据的精确度必然很低,此时,样本含量越大,提供的错误信息就越多,劳民伤财,的确毫无价值;样本含量过小,也不妥当。因为人们研究的问题往往比较复杂,观测的指标常带有变异性。要想反映带有变异性的指标

(常称其为随机变量)的变化规律性,必须在相同实验条件下进行多次独立的重复实验,才有可能使随机变量的规律性真实地显露出来。

样本含量通常需要区分,是整个实验研究中的样本含量,还是实验研究中某特定小组中的样本含量。在多因素实验研究中,千万不要仅看  $N$  的数值,而应着重考察各组的  $n$  数值。

总而言之,要想使实验结果达到一定的精确度,又不浪费人力、物力和时间,需要根据具体的情况、基本信息和精确度要求,估计所需的各小组的最小样本含量。这就是通常所讲的样本含量估计问题。

### 5 如何确定合适的样本含量

一般来说,样本含量越大,即重复的次数越多,就越能反映出客观事物的统计规律性及真实情况。但样本含量并非多多益善。因为样本过大,既造成不必要的浪费,又延长研究周期,所以不可一概排斥小样本。所谓样本含量估计,就是要求在保证研究结论具有一定可靠性的条件下,确定最少的实验单位数。样本含量越大或重复次数越多,越能反映变异的客观真实情况。但若认为样本含量过大,即使无专业意义的很小差异,也可得到有统计学意义的结论,实则浪费人力物力。若样本含量过小,即使有专业意义的差异,也可能没有统计学意义,造成假阴性错误。因此,正确估计样本含量十分重要<sup>[11]</sup>。

这是一个无法准确回答的问题,估计合适的样本含量,要涉及下列有关内容<sup>[12]</sup>:①拟开展的研究类型。通常分为调查研究、实验研究或临床试验研究。②研究中将涉及到的因素的个数及其水平数的多少。在单组设计、配对设计、成组设计中,估计样本大小问题相对比较简单,有现成的公式可用;而在单因素多水平设计中,虽然也有公式,但问题就要复杂多了,而在多因素析因设计和其他多因素设计中,样本大小的估计就更复杂了。③观测的效应指标的性质。若效应指标是定量的(如比较平均值),则所需样本含量较小;若效应指标是定性的(如比较率或比例),则所需样本含量较大。④必不可少的前提条件。一般来说,在估计样本含量时,研究者需提出某些精度要求,并提供某些先验知识,方可进行估计。所谓精度要求,通常指事先规定允许犯 I 型(或假阳性)错误的概率  $\alpha$ 、检验功效或把握度  $1-\beta$ [(这里  $\beta$  为允许犯 II 型(或假阴性)错误的概率)]。 $\alpha$  定得越小,  $1-\beta$  定得越大,所需样本含量就越大。所谓先验知识,就是根据专业知识、文献资料或预实验结果获得的由样本推断总体的一些信息,如:已知  $\delta$ (最大容许误差或差值)、 $\sigma$ (标准差)等。⑤研究结果将来的应用范围。假定要制定临床上某定量指标的正常值范围,若此正常值范围仅用于一个较小的人群(如某城市),则所需要的样本含量就可以小一些;若此正常值范围将用于一个较大的人群(如全国),则所需要的样本含量就相当大;否则,其代表性就很差。

总之,要想确定合适的样本含量,应综合考虑以上几个方面,找到现成的估计样本含量的计算公式或表格或计算机软件,获得有根据的样本含量的估计值。估计样本含量的方法有 3 种:经济法、计算法和查表法<sup>[13]</sup>。各种实验设计条件下样本

含量的估计是一个比较复杂的问题,有待统计工作者不断探索和研究。

在样本量方面常见的错误有:有时盲目追求大样本,导致质量差效率低;有时又严重忽视重复实验的作用,在不同实验条件下,仅做一次实验,因为生物体的个体差异很大,在相同条件下重复实验次数过少,随机变量的规律性无法正确地显露出来;更有甚者,用“个案”上得出的结论取代一般规律<sup>[14]</sup>。

### 6 临床研究实施重复原则的常用措施<sup>[15]</sup>

严格按照研究的目的规定研究对象的性质与范围。保证临床研究有足够的样本含量。实验设计是科学研究计划内关于研究方法步骤的一项内容。科学研究结果的价值首先决定于研究目的,而研究结果是通过一定方法步骤来获得的。一个周密的实验设计,能合理地安排各项实验因素,正确地估计样本含量大小,严格地控制实验误差,从而用较少的人力、物力、时间,最大限度地获得丰富而可靠的资料。反之,如果实验设计存在着缺点,就可能造成不应有的浪费,且足以减损研究结果的价值。因此,在制定研究计划时,认真考虑实验设计是必要的。

### 参考文献

- [1] 刘桂芬. 医学统计学. 北京: 中国协和医科大学出版社, 2007. 165-170.
- [2] 杨甫德. 设计药物临床试验的主要原则. 中国临床康复, 10(44): 127-129.
- [3] 胡良平. 医学统计实用手册. 北京: 人民卫生出版社, 2004. 12-13.
- [4] 胡良平. 医学统计实战练习. 北京: 军事医学科学出版社, 2007. 4.
- [5] 同志超, 杨镇, 同志勤, 等. 骨髓间质干细胞体外定向诱导分化为软骨细胞的实验研究. 中国骨伤, 2008, 21(5): 362-364.
- [6] 彭磊, 胡蕴玉, 徐华梓, 等. 间歇性压力培养环境对兔骨髓基质干细胞增殖的影响. 中国骨伤, 2007, 20(2): 92-93.
- [7] 苏佳灿, 管华鹏, 张春才, 等. 冲击载荷作用下骨盆三维有限元分析及其生物力学意义. 中国骨伤, 2007, 20(7): 455-457.
- [8] 苏友新, 郑良朴, 陈智能, 等. 强骨宝方对外培养成骨细胞影响的实验研究. 中国骨伤, 2007, 20(6): 394-396.
- [9] 周祖彬, 曾炳芳. 三角纤维软骨复合体解剖及生物力学研究. 中国骨伤, 2006, 19(11): 666-667.
- [10] 孙治安, 张秀清. 医学科研设计中对照、随机和重复原则的常见错误. 河北医药, 2000, 6(2): 159-160.
- [11] 徐天和, 王玖. 医学实验设计: 第三讲重复原则. 中国医刊, 2005, 40(9): 56-57.
- [12] 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. 33-38.
- [13] 范彬, 吴基良, 余英宏. 医学统计与科研设计. 北京: 科学出版社, 2003. 105-109.
- [14] 胡良平. 科研课题的研究设计与统计分析. 北京: 军事医学科学出版社, 2008. 19.
- [15] 罗家洪, 薛茜. 医学统计学. 北京: 科学出版社, 2008. 232-233.

(收稿日期: 2009-07-28 本文编辑: 李为农)