

· 继续教育园地 ·

《中国骨伤》杂志论文中存在的统计学错误辨析(3)

葛毅, 胡良平, 周诗国, 柳伟伟

(军事医学科学院生物医学统计学咨询中心, 北京 100850)

本文对《中国骨伤》杂志 2006 年与 2007 年的第 3 期所载临床研究和基础研究类文章进行了仔细阅读, 针对论文中统计学方法的应用情况进行了考察, 就实验设计和统计分析两个方面存在的问题进行了辨析, 提出改进的建议和意见, 希望能为读者在提高科研工作科学性和严谨性方面提供一定的帮助。

1 实验设计方面典型错误案例辨析

在骨科的临床研究中, 创新或者改进的治疗方法是研究的一个重要方向, 而评价一种治疗方法的好坏重要的是要有一个比较基础, 也就是要有对照, 此外使重要非实验因素在组间均衡一致也是一个十分重要的方面。下面分别举例说明。

例 1, 《等离子体髓核成形术的疗效观察》^[1]一文中, 研究目的为探讨髓核成形术治疗腰椎间盘突出症的作用机制并评价其疗效。传统开放手术治疗腰椎间盘突出症需经后路手术切除椎板, 摘除髓核。近年来出现了一系列治疗腰椎间盘突出症的微创介入方法, 如胶原酶髓核溶解术、经皮髓核摘除术等。该科于 2002 年初引进美国 ArthroCare 等离子体髓核成形术治疗腰椎间盘突出症。

此研究选择 64 例单节段腰椎间盘突出症患者, 男 37 例, 女 27 例; 年龄 26~60 岁, 平均 42.9 岁; 病程 1~30 年, 平均为 4.2 年。L_{4,5} 突出 30 例, L₅S₁ 段突出 34 例, 根据椎间盘造影分为 I 型 39 例, II 型 25 例。于髓核成形术前后测定椎间盘内压, 观察其变化, 同时行椎间盘造影, 判断椎间盘突出类型, 并于术后 6 个月评价其疗效。统计分析实验数据经 SPSS 10.0 统计软件进行分析。髓核成形术前两型椎间盘内压的比较采用成组 *t* 检验; 两型患者手术前后椎间盘内压值的变化采用配对 *t* 检验; 椎间盘内压变化与疗效相关性分析采用等级资料的秩相关分析方法。结果: I 型患者疗效: 优 23 例, 良 12 例, 可

4 例, 差 0 例, 优良率 89.74%。II 型患者疗效: 优 7 例, 良 9 例, 可 5 例, 差 4 例, 优良率 64%。根据造影结果将病变椎间盘分为 2 型: I 型为突出型, II 型为破裂型。在髓核成形术前, I 型患者椎间盘内压明显高于 II 型患者 ($P < 0.01$); 髓核成形术后, I 型患者椎间盘内压较术前明显下降 ($P < 0.01$), II 型患者椎间盘内压较术前变化不明显 ($P > 0.05$); I 型患者术后疗效优于 II 型患者 ($P < 0.05$)。结论: 作为一种新技术, 髓核成形术能明显降低 I 型患者椎间盘内压, 改善其症状, 是一种安全、有效的微创介入手术。

辨析: 此研究的目的是为探讨髓核成形术治疗腰椎间盘突出症的作用机制并评价其疗效, 并得出髓核成形术是一种安全、有效的微创介入手术。但是该研究只是对这一种方法进行了研究, 评价优良率, 而并没有与目前其他同类方法进行比较, 因此在实验设计上犯了缺乏对照的错误, 结论为该方法有效只是一家之言, 并没有参照。并且在病例的选择上, 年龄跨度过大, 青年人 (20 岁) 和老年人 (60 岁) 毕竟治疗效果是不同的, 病程 (1~30 年) 也对疗效有影响, 仅有 1 个治疗组, 无法排除这 2 个重要非实验因素对观测结果的影响, 使结论的可信度大打折扣。

在统计分析方法上采用 *t* 检验进行 2 型治疗前后椎内压的比较也是不妥的, 它割裂了实验的整体性, 容易增大犯假阳性错误的概率。由于该研究的实验因素为分型 (I, II 型 2 个水平)、时间 (术前, 术后 2 个水平) 2 个因素 (见表 1), 而每一个患者都在治疗前和治疗后分别被测椎内压, 因此应采用具有一个重复测量的两因素设计定量资料方差分析处理为宜。

原作者评价两型 (I 型, II 型) 患者的疗效时未采用任何统计分析方法, 只是直观地比较了优良率便下结论, 这样使得结

- 37:218-227.
- 12 Pasapula C, Mandalia V, Aslam N. The floating shoulder. Acta Orthop Belg, 2004, 70:393-400.
- 13 Hashiguchi H, Ito H. Clinical outcome of the treatment of floating shoulder by osteosynthesis for clavicular fracture alone. J Shoulder Elbow Surg, 2003, 12:589-591.
- 14 Ledger M, Leeks N, Ackland T, et al. Short malunions of the clavicle; an anatomic and functional study. J Shoulder Elbow Surg, 2005, 14:349-354.
- 15 Bozkurt M, Can F, Kirdemir V, et al. Conservative treatment of scapular neck fracture; the effect of stability and glenopolar angle on clinical outcome. Injury, 2005, 36:1176-1181.
- 16 Hagino T, Ono T, Hamada Y. Unusual double clavicle fracture com-

plicated by ipsilateral scapular neck fracture. J Orthop Sci, 2002, 7: 417-419.

- 17 Oh W, Jeon IH, Kyung S, et al. The treatment of double disruption of the superior shoulder suspensory complex. Int Orthop, 2002, 26:145-149.
- 18 Debeer P, Robyns F. Fracture of the scapular spine in a patient with a Delta III prosthesis. Acta Orthop Belg, 2005, 71:612-624.
- 19 de Beer JF, Berghs BM, van Rooyen KS, et al. Displaced scapular neck fracture; a case report. J Shoulder Elbow Surg, 2004, 13:123-125.
- 20 Esenkaya I. Surgical treatment of scapular fractures. Acta Orthop Traumatol Turc, 2003, 37:33-40.

(收稿日期: 2007-12-27 本文编辑: 王玉蔓)

表 1 I、II 型椎间盘患者治疗前后的椎间盘内压 ($\bar{x} \pm s$)

造影分型	例数	椎间盘内压(CmH ₂ O)	
		时间: 术前	术后
I 型	39	10.68±0.95	2.28±0.76
II 型	25	2.13±0.51	2.09±0.49

论的可信度很低,因为没有用统计分析方法进行分析,便不知这样的结果是抽样误差造成的还是确实如此。

正确的统计分析方法如下:因为此处的效应指标为“疗效”,而且有“优、良、可、差”4 个水平,所以效应指标为等级变量(或有序变量)。因此,采用结果变量为有序变量的单向有序 2×4 列联表(见表 2)资料的秩和检验或 Ridit 分析(各型例数大于 100 时用更合适)处理,由于本例样本量较小,因此应采用秩和检验分析(本例,近似 $\chi^2=8.383, P=0.0038$),这样才能比较在该手术方法下,2 型患者的治疗效果究竟如何(本例结论为 I 型患者的疗效优于 II 型患者的疗效)。

表 2 I、II 型椎间盘患者治疗效果

造影分型	例数			
	疗效: 优	良	可	差
I 型	23	12	4	0
II 型	7	9	5	4

以上是在实验设计上有错误的例子,例 2 与例 1 在临床研究方向上类似,但其从实验设计到统计分析都是比较合理的。

例 2,《肱骨髁上截骨术后可吸收与金属张力带固定治疗肘内翻的研究》^[2]一文中,肘内翻是儿童肱骨髁上骨折常见的远期并发症,肱骨髁上楔形截骨矫形术是传统治疗方法,术后多采用金属螺钉或克氏针钢丝张力带固定,需 II 期手术取出。该研究目的为比较肱骨髁上楔形截骨术后可吸收张力带(可吸收螺钉或棒和可吸收线)与金属张力带(克氏针和钢丝)固定治疗肘内翻的疗效。方法:采用 2 种张力带固定肱骨髁上楔形截骨术后骨端治疗 33 例肘内翻。可吸收张力带固定组 15 例,其中男 12 例,女 3 例;年龄 13~24 岁,平均 16 岁;术前内翻角为 17°~34°,平均 26°。金属张力带固定组 18 例,其中男 15 例,女 3 例;年龄 12~22 岁,平均 14 岁;术前内翻角为 17°~40°,平均 29°。随访时间为 18~69 个月,并根据术后畸形矫正和关节功能恢复情况评定疗效,结果为可吸收张力带固定组:优 13 例,良 2 例,差 0 例;金属张力带固定组:优 15 例,良 2 例,差 1 例。经运用 SAS 9 统计软件,采用非参数法(Wilcoxon)比较 2 组疗效,其结果为 $P=0.7495>0.05$,2 种治疗方法临床疗效比较无统计学差异。结论:可吸收张力带是固定肱骨髁上截骨治疗肘内翻的有效方法之一,具有无须 II 期手术取出内固定物的优点,简单方便,降低感染机会。

辨析:从本例可以看出该项研究的主要目的是探讨肱骨髁上楔形截骨术后可吸收张力带(可吸收螺钉或棒和可吸收线)的治疗效果及优势。为了说明可吸收张力带治疗方法的优点,该研究采用了与常用金属螺钉或克氏针钢丝张力带治疗方法进行了比较,在疗效没有统计学差别的前提下,得出了可吸收张力带治疗方法无须 II 期手术的优点。在病例的选取上也对治疗效果有影响的年龄、严重程度加以均衡。选取的各组年龄

分别在 13~24 岁和 12~22 岁的青少年,术前内翻角分别为 17°~34°和 17°~40°,2 组内翻角度相差不多,在性别构成比上也较均衡,所以 2 组的病例可比性较强,结论有说服力。在统计分析方法选择上也是合理的,因为此研究中两组疗效的比较是一个结果变量为有序变量的单向有序 2×3 列联表(原作者未给出表格),编制出表格后见表 3,所以采用 Wilcoxon 秩和检验分析此资料是合适的。

表 3 采用两种张力带固定治疗效果比较

张力带类型	例数		
	疗效: 优	良	差
可吸收	13	2	0
金属	15	2	1

由以上 2 个例子可以看出对一种治疗方法的优劣性的评价在骨伤治疗中是比较常见的,通常都采用对治疗方法的疗效分析,但是为了使结论能有说服力,应该在实验设计时设立对照,没有对照就没有比较的基础,就不能对疗法的疗效做出一个准确的评价,因而也就成为“老王卖瓜,自卖自夸”,失去了研究的意义。同时还要均衡一些对疗效有影响的重要非实验因素,如严重程度、年龄、性别等,使得各组之间有可比性。在进行各组间疗效的比较上,由于资料通常为结果变量为有序变量的单向有序的二维列联表资料,即表中仅结果变量的取值为有序的,而原因变量是无序的,如某资料中,原因是“治疗与对照”、结果是“治愈、显效、好转、无效”,或某指标的取值为“-、+、++、+++”等,如表 2 和表 3。如果仅原因变量为有序的而结果变量是名义的,则通常情况下仍将其视为“双向无序的 R×C 列联表”资料。

结果变量为有序变量的单向有序 R×C 列联表资料的统计分析方法可选用秩和检验、Ridit 分析以及有序变量的 logistic 回归分析。如例 2 采用 Wilcoxon 秩和检验分析是正确的。可能有不少人一看形如表 2 和表 3 的资料便采用 χ^2 检验分析,这是不正确的,因为 χ^2 检验虽常用于定性资料的分析,但与变量的有序性没有任何联系,用一般的 χ^2 检验进行分析,得到的结论是 2 组患者在全部疗效等级上的频数分布是否相同,而不能得出 2 组疗效之间的差别是否具有统计学意义的结论。

2 统计分析方面典型错误案例辨析

统计分析方法的正确运用是进行资料分析的关键,这不仅有利于计算结果的解释,更重要的是提高结论可信度的又一个关键环节。选择正确的统计分析方法要根据研究目的、资料性质和具备的前提条件以及资料所对应的实验设计类型决定。

例 3,《田胡止痛颗粒对大鼠椎间盘磷脂酶 A₂ 及神经根炎性细胞的影响》^[3]一文,实验目的为应用自体椎间盘移植的动物模型探讨田胡止痛颗粒治疗腰椎间盘突出症的相关机制。方法:72 只大鼠随机分为假手术组(SG)、模型对照组(CG)、芬必得组(FG)、田胡止痛颗粒组(TG),每组 18 只,采用自体尾椎椎间盘植入 L₅ 神经根背侧的方法制备大鼠腰椎间盘突出模型。各组大鼠于术后 1 d 开始灌胃,SG 和 CG 用蒸馏水,FG 和 TG 分别应用芬必得和田胡止痛颗粒

表 4 不同药物在不同时间段的 PLA₂ 和炎性细胞浸润数目的比较 ($\bar{x} \pm s$)

药物种类	PLA ₂ 活性(U)			炎性细胞数(个数/HP)			
	时间:	7 d	14 d	28 d	7 d	14 d	28 d
SG		3.968±0.186	4.120±0.414	3.917±0.258	7.50±1.049	4.67±1.033	3.33±0.816
CG		9.845±0.218	8.903±0.389	9.131±0.262	18.67±2.066	16.17±1.169	15.50±1.049
FG		5.797±0.261	5.301±0.295	4.818±0.265	9.367±2.160	9.33±4.033	5.00±1.414
TG		8.293±0.765	5.452±0.347	4.973±0.484	14.67±1.211	10.33±1.751	5.50±1.049

的悬浊液,并分别于术后的 7、14、28 d 处死动物,检测原移植椎间盘(SG 取残留尾椎的椎间盘组织)磷脂酶 A₂ 的活性,对 L₅ 神经根进行病理切片检查和炎性细胞计数。并采用 SPSS 13.0 统计软件,组间应用单因素 4 水平方差分析中的 LSD 法或 Dunnett's 法(方差不齐)进行两两比较。结果:TG 椎间盘磷脂酶 A₂ 的活性在用药 7 d 后与 CG 相比有统计学差异($P < 0.05$),14 及 28 d 后比较有统计学差异等结果(见表 4)。结论:田胡止痛颗粒能够降低大鼠突出椎间盘磷脂酶 A₂ 的活性,减轻神经根的炎性反应,并且起效快,促进神经功能的恢复。

辨析:此研究的目的是应用自体椎间盘移植的动物模型探讨田胡止痛颗粒治疗腰椎间盘突出症的相关机制。从全文可以看出主要是研究使用田胡止痛颗粒磷脂酶 A₂ 的活性和炎性细胞数的变化。本实验将 72 只大鼠随机分为 4 个组,其中假手术组(SG 组)未建立模型,而其余 3 组都经手术建立了模型,其中 FG、TG 两组在建模的基础上分别使用了田胡止痛颗粒和芬必得;SG 组用蒸馏水灌胃。所以上表是该研究的表现型^[4]。这种资料的实验设计类型常被误认为“单因素 4 水平设计”,进而选择单因素 4 水平设计定量资料的方差分析处理所收集的定量资料。但实际上本实验包含 3 个实验因素,即“是否建立模型”、“在模型基础上药物种类(田胡止痛颗粒、芬必得)”和“术后时间(7 d、14 d、28 d)”。在已设计的 4 个组中,没有将上述诸因素各水平进行全面组合,而只是实施了其中的一部分,所以,本例属于多因素非平衡组合实验,不是一个标准的多因素设计。

对于这种资料,需要结合专业知识,将“组别”拆分成几个不同的组合。拆分的依据是所形成的每一种组合在专业上能解释得通,在统计学上又有相应的标准设计与之相对应。拆分结果有以下 2 种情况。

组合 1:SG 组与 CG 组;建模与否对应的检测指标平均值之间的差别是否有统计学意义。

组合 2:CG 组、FG 组与 TG 组;建模基础上用不同药物后,对应的检测指标平均值之间的差别是否有统计学意义。

由于各组都是在 3 个不同的时点处死动物,所以在上 2 种组合下还有一个时间因素,在不知分组因素和时间因素对观测结果影响是否有主次之分的前提下,上 2 种组合都可视为两因素析因设计,组合 1 为 2×3 析因设计,组合 2 为 3×3 析因设计,应采用相应设计定量资料方差分析。因此原作者将这些组放在一起比较,并用单因素 4 水平方差分析进行分析是不正确的,不仅不能对每个因素的效应大小作出合理评价,而且还无法考查“时间”因素与“组别”两因素之间的交互作用。这 2 个组合分别列为表 5 和表 6(标准型)。原作者在进行组间比较时,应用单因素 4 水平设计定量资料方差分析中的 LSD 法或 Dunnett's 法进行两两比较。由于本实验进行两因素析因设计定量资料方差分析和单因素 4 水平设计定量资料方差分析的组数不同,从而误差项的自由度也不同,所以有可能导致结论的不一致。此外,原作者只对模型组与假手术组进行比较,而两个用药组间没有比较,因此两种药物的作用是否相同则不得而知了。

例 4,《循环静水压力对维持兔软骨细胞表型的影响》^[5]一文中,研究目的为不同频率和大小的静水压对兔软骨细胞表型的影响。取 4 只新西兰雄性白兔进行软骨细胞分离和培养,并将培养的细胞分成 3 份。A 组循环静水加压,每天加压 4 h;B 组持续静水加压,每天加压 24 h;C 组常压对照组。各组选 1、4、7 代细胞都分别接种在 96 孔板上,每代接种 10 孔,接种密度同培养瓶,分为 A、B、C 3 组。分别在接种贴壁后第 1~7 天终止培养,并用酶标仪测每孔的吸光度(OD 值),采用 SPSS 10.0 统计分析软件运用单因素 3 水平方差分析方法对数据进行统计处理,分别看各代在不同时间点的 3 个组间是否有差别,并以时间为横坐标,OD 值为纵坐标绘制生长曲线图。结论为一定的循环静水压力能维持软骨细胞的合成分泌功能,

表 5 模型组与假手术组不同时间段的 PLA₂ 和炎性细胞浸润数目的比较 ($\bar{x} \pm s$)

是否建模	PLA ₂ 活性(U)			炎性细胞数(个数/HP)			
	时间:	7 d	14 d	28 d	7 d	14 d	28 d
否		3.968±0.186	4.120±0.414	3.917±0.258	7.50±1.049	4.67±1.033	3.33±0.816
是		9.845±0.218	8.903±0.389	9.131±0.262	18.67±2.066	16.17±1.169	15.50±1.049

表 6 不同药物在不同时间段的 PLA₂ 和炎性细胞浸润数目的比较 ($\bar{x} \pm s$)

药物种类	PLA ₂ 活性(U)			炎性细胞数(个数/HP)			
	时间:	7 d	14 d	28 d	7 d	14 d	28 d
CG		9.845±0.218	8.903±0.389	9.131±0.262	18.67±2.066	16.17±1.169	15.50±1.049
FG		5.797±0.261	5.301±0.295	4.818±0.265	9.367±2.160	9.33±4.033	5.00±1.414
TG		8.293±0.765	5.452±0.347	4.973±0.484	14.67±1.211	10.33±1.751	5.50±1.049

表 7 兔软骨细胞在不同压力、代次条件下 OD 值的变化($\bar{x} \pm s$)

编号	OD 值									
	分组:	A 组			B 组			C 组		
	代次:	1 代	4 代	7 代	1 代	4 代	7 代	1 代	4 代	7 代
	时间:	1~7 d	1~7 d	1~7 d	1~7 d	1~7 d	1~7 d	1~7 d	1~7 d	1~7 d
1	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
2	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
3	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
4	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
5	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
6	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
7	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
8	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
9	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	
10	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	X...X	

注：“1~7 d”为 7 个时间点，分别为第 1 天、第 2 天、...、第 7 天有助于软骨细胞特有表型的稳定。

辨析:该研究的目的是为不同频率和大小的静水压对兔软骨细胞表型的影响。将兔软骨细胞按不同条件分成 3 个组，观察不同条件、不同子代、不同时间的细胞生长情况，并在不同的组合条件下采用单因素 3 水平设计定量资料方差分析，来分析细胞 OD 值的变化。这只是该研究的表现型，从全文分析，其原型应为考虑“不同压力条件”、“代次”、“时间”3 个因素对兔软骨细胞 OD 值的影响。因此将这 3 个因素分开，采用单因素分析方法进行分析是不对的。由此可以根据该研究的原型(表达研究的真实目的)，将这 3 个因素综合考虑，得出其标准形如表 7 所示。

由于 3 个组的样本是 4 只新西兰雄性白兔进行软骨细胞分离和培养并分成的 3 份，所以它是一个重复测量因素，“代次”是这一批细胞培养出的不同代，所以也是一个重复测量因素，“时间”是对同一批细胞在不同时间点上的观测，所以也是一个重复测量因素。由此，这是具有 3 个重复测量的 3 因素设计，但是这还不是一般的重复测量设计，因为“代次”与“时间”之间还有嵌套关系，应该根据这 3 个因素之间的关系建立相应的统计模型进行分析。原作者使用了单因素 3 水平设计定量资料方差分析只是根据该研究的表现型，因为重复测量数据之间具有一定的相关性，按简单的单因素分析处理，不但不能反映出数据间的相关性，而且割裂了因素之间的交互作用和嵌套关系。如果要考查不同压力条件下的 OD 值随时间推移的变化规律，还应采用正交多项式回归分析，这将会更客观地反映 OD 值的变化趋势。

由以上 2 个例子可以看出，在选用统计分析方法时应从资料的性质加以分析，通过表现型看清其原型，并分析出对应的标准型后才能正确选用统计分析方法。具体地说，在判断资料性质时，要看从每一个受试对象身上测得指标的具体内容是什么，而不要仅从资料的表面现象去判断资料性质，要通过表面现象去看资料及研究的实质。在判断资料的实验设计类型时，应确定资料所涉及的因素的个数和各因素水平的组合情况，还要注意研究者在收集和整理资料时所采用的方法，综合分析，才能对实验设计类型做出较为准确的判断。

综上所述，实验设计与统计分析在医学科研工作中是相互联系的，实验设计的正确与否直接关系到统计分析方法的选择，因此实验设计与统计分析应贯穿于整个科研工作的始终，而不是仅仅局限于最后的实验数据处理(分析)阶段。

参考文献

- 1 蒋林, 陈庆. 离子体髓核成形术的疗效观察. 中国骨伤, 2006, 19(3): 134-136.
- 2 陈爱民, 陈梓锋, 叶艳平, 等. 肱骨髁上截骨术后可吸收与金属张力带固定治疗肘内翻的研究. 中国骨伤, 2007, 20(3): 167-169.
- 3 翁凤泉, 康乐, 黄经光, 等. 田胡止痛颗粒对大鼠椎间盘磷脂酶 A₂ 及神经根炎性细胞的影响. 中国骨伤, 2007, 20(3): 176-179.
- 4 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. 20-43.
- 5 黄正, 傅文斌, 冯伟, 等. 循环静水压力对维持兔软骨细胞表型的影响. 中国骨伤, 2007, 20(3): 180-183.

(收稿日期: 2008-02-29 本文编辑: 李为农)