

· 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(8)

柳伟伟, 胡良平

(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, 2008, 21(9): 713-715 www.zggszz.com

本文在前 7 期统计学错误辨析系列文章的基础上, 继续对《中国骨伤》杂志 2006 年与 2007 年第 8 期所刊载的临床研究、基础研究和经验交流文章进行考察。由于之前的 7 篇文章中对于统计学的各个方面已经有了较为全面的说明, 这里不再作整体性的讨论, 而只是针对容易出现的一些问题给予强调, 主要涵盖实验设计中的对照原则、重复测量设计的统计分析和不同类型列联表的分析。

1 实验设计中的对照原则及错误辨析

在前述两期杂志的论文中, 研究者普遍都能够重视随机原则和均衡原则, 在分组时进行随机化, 控制重要的非实验因素在实验因素的水平之间达到均衡一致, 在临床试验中, 对于不同组别间治疗前的基线数据进行假设检验。相对来讲, 在严格遵守对照原则方面问题显得稍微突出一些, 在有些文章中, 没有设置必要的对照组, 本文将结合实例展开这方面的讨论。

所谓对照, 是指在实验研究中, 要想得出什么处理或治疗方法效果好, 必须交代“与谁比较”, 比较至少要在两个事物之间进行, 在统计学上通常是在两个组之间进行, 其中一个组被称为实验组, 另一个组被称为对照组。对照的作用在于建立比较的基础, 提高鉴别的能力和结论的说服力, 换言之, 缺乏对照的研究是没有说服力的。当然, 对照不全或对照设置得不合理也是没有说服力的。对照的形式有多种, 包括空白对照、标准对照、相互对照、实验对照、历史或中外对照。违反对照原则的常见错误类型主要有缺乏对照、对照不全、对照过剩以及假对照等^[1-2]。

例 1: 某临床研究^[3]选择 35 例颈椎病患者, 男 23 例, 女 12 例; 年龄 47~68 岁, 平均 56.4 岁。34 例脊髓型颈椎病, 1 例为颈椎术后翻修, 椎管矢状径 4.6~9.3 mm; 椎管矢状面狭窄率达 39%~84%, 平均 49.3%; 减压节段 C₃-C₅ 9 例, C₃-C₆ 20 例, C₄-C₇ 6 例。采用 Cervifix 固定, 垂直水平悬吊保持开门状态治疗。结果优 28 例, 良 4 例, 差 3 例。JOA 评分术前平均 8.0 分, 术后平均 15.7 分。结论为 Cervifix 固定垂直水平悬吊改良单开门椎管成形术操作简单、安全, 维持开门效果好, 术后患者颈部症状恢复快、效果满意, 是治疗多节段颈椎病及颈椎管狭窄症的一种有效可行的方法。

辨析: 此研究的目的是评价 Cervifix 固定垂直水平悬吊改良单开门椎管成形术的疗效, 其疗效评定标准是按 JOA 评分方法计算出患者术后改善率, 然后按照 JOA 评分术后改善率将疗效分为优、良、可、差 4 个等级。目前应用最多的术式是

采用粗丝线通过棘突根部的预穿孔将开门后椎板缝合悬吊于同节段的小关节囊和周围韧带上, 术后围颈制动 1~3 个月, 这种方法导致许多患者存在术后颈部长期疼痛等问题。文章中提出的方法就是为了解决这些问题, 所以, 该研究应该将应用最多的术式设置为对照, 这里违反了对照原则, 错误类型为缺乏对照。在未设置对照、没有进行假设检验的情况下, 对此法做出“简单、安全, 维持开门效果好, 术后患者颈部症状恢复快、效果满意”的结论是没有说服力的。正确的做法是设立合理的对照, 进行随访观察获得试验数据, 通过恰当的统计分析方法对两种手术方法的疗效进行比较, 这样得到的结论才是可信的。

例 2: 某临床研究^[4]中选取多节段脊髓型颈椎病 38 例, 男 29 例, 女 9 例; 年龄 35~78 岁, 平均 56.5 岁; 慢性发病 18 例, 急性发病 9 例, 轻度外伤或劳累后发病 11 例; 病程 15 d~16 年, 平均 31 个月; 病变累及 2 个节段 19 例, 3 个节段 14 例, 4 个节段 5 例; 曾行后路全椎板减压术 5 例, 单开门 1 例。采用保留颈椎椎体后侧壁植骨融合的方法治疗以提高颈椎椎体结构的相对稳定及植骨融合率, 减少并发症的发生。结果优 19 例, 良 14 例, 可 4 例, 差 1 例。

辨析: 该研究的目的是探讨经颈前路植骨融合治疗多节段颈椎病的临床疗效。研究者也没有设置必要的对照组, 没有进行统计分析, 只是根据采用保留颈椎椎体后侧壁植骨融合方法治疗的一组患者的治疗结果就得出最后的结论。可以考虑采用标准或公认的某种疗法作为对照, 进行两组间的比较, 进而对疗效作出科学的判定。同时也要考虑到年龄和病程等可能对结果产生影响的因素, 在分组时最好按这些因素采用分层随机化。

2 重复测量资料的统计分析及错误辨析

重复测量就是同一个受试对象(如个体、样品等)在接受某种处理后, 即在某个因素不同水平条件下或某几个因素水平组合条件下被重复观测 2 次或 2 次以上的实验研究问题。重复测量资料可以是定量的, 也可以是定性的。这种资料有其特殊性, 对同一个个体而言, 多次重复观测得到的结果数据之间往往存在一定的相关性, 而且对于不同的资料, 相关关系存在多种不同的形式, 其中比较常见的是相隔越近的数据之间相关性越大, 相隔越远的数据之间的相关性越小。正是由于这种相关性的存在, 重复测量设计定量资料很可能不满足一般的方差分析模型的一些假定, 这时仍然使用 *t* 检验或常规的方差分析, 得到的结果是不可信的, 此时应当使用重复测量设计

定量资料的方差分析。当重复测量资料满足球形对称或复合对称性假定时,可以使用 SAS 软件中的 GLM 过程,即一般线性模型过程分析;如果资料不满足上述假定,则可以使用混合模型进行分析,并且可以定义不同的相关形式,也就是协方差结构,混合模型由 MIXED 过程完成。

重复测量定性资料的分析方法有加权最小二乘法(weighted least square, WLS)、广义估计方程(generalized estimating equations, GEE)、广义线性混合模型(generalized linear mixed models, GLMM) 和非线性混合模型 (nonlinear mixed model)。在 SAS 软件中,可以分别用 CATMOD 过程、GENMOD 过程、GLIMMIX 过程和 NLMIXED 过程实现^[5]。

例 3:某基础研究^[6]目的为观察补阳还五汤联合骨髓间质干细胞(MSCs)移植对大鼠脊髓损伤后神经功能恢复以及移植的 MSCs 迁移情况的影响,并探讨其作用机制。取 80 只 SD 大鼠,其中 70 只用改良 Allen 法制备大鼠 T₁₀ 脊髓损伤模型,随机分为中药+MSCs 组 20 只、MSCs 组 20 只、假手术组(无脊髓损伤)10 只、中药组 20 只、空白对照组(无治疗)10 只。各组大鼠于术后 1、3、5 周观察神经功能恢复、免疫组化检测带标记的 MSCs 迁移情况。在统计学分析部分,原文的表述为“行单因素方差分析 F 检验”。研究中斜板试验的结果见表 1。

表 1 各组大鼠损伤后斜板试验结果($\bar{x} \pm s, ^\circ$)

Groups	Angle of incline		
	1 weeks	3 weeks	5 weeks
Drugs+MSCs	32.250±5.399*	37.170±3.839*	41.500±5.762* [△]
MSCs	30.700±3.840*	34.220±3.264*	33.810±5.406*
Buyang-Huanwu	29.550±3.927*	32.220±4.181*	32.250±5.508*
Blank control	21.900±2.234	20.500±3.071	20.000±4.050

注:与 MSCs 组比较, [△]P<0.05;与空白对照组比较, *P<0.01

辨析:本研究共设计了 5 个独立的组,但这 5 个组并不是一个标准的多因素设计,而是属于多因素非平衡组合实验,关于具体的拆分方法,由于前 7 期中已有详细的介绍,此处不再赘述。从 5 个组中拿出其中的 4 个组(中药+MSCs 组、MSCs 组、中药组、空白对照组)与时间因素同时考虑,也就是表 1 中所列的全部内容,属于具有一个重复测量的三因素设计定量资料,这里是否使用中药、是否移植 MSCs 为两个实验因素,各有两个水平;时间是重复测量因素,有 3 个水平。同一只大鼠在不同时间点的斜板试验结果可能存在一定的相关性。原作者使用单因素方差分析,显然是将上述 4 组看作同一因素的不同水平,然后在不同的时间点上分别进行比较,这是不正确的。对于该资料,应该使用具有一个重复测量三因素设计定量资料的方差分析。如果需要不同组别之间的相互比较,可以在方差分析的基础上进一步做两两比较。另外,“行单因

素方差分析 F 检验”的表述也有误,正确的表达应该是“采用具有一个重复测量三因素设计定量资料的方差分析”。为使实验设计类型判断不易出错,应将表 1 中“Groups”列改为 2 列,其表头分别为“是否使用中药”与“是否移植 MSCs”,感兴趣的读者可试着去重新编制此表,以期达到不言自明之效果。

例 4:某项基础研究^[7]中想探讨发育期髋关节股骨头骨骺关节软骨缺血再灌注后的诱导型一氧化氮合成酶(iNOS)及热休克蛋白 70(HSP70)的表达,制作 4 周龄 SD 大鼠缺血再灌注及对照模型,每组各 40 只,于术后 3、6、12、24、48 h,5 d,2、4 周不同时间点股骨头骨骺关节软骨 iNOS 及 HSP70 进行免疫组织化学检测。统计学分析采用组间方差分析及配对 t 检验,P<0.05,差异有显著性意义。其中缺血再灌注组及对照组各时点股骨头骨骺关节软骨浅层、中层的 HSP70 表达灰度见表 2。缺血再灌注组和对照组各时点关节软骨浅层及中层 HSP70 表达灰度经配对 t 检验的比较,在 24 h 至 5d 时点 P<0.05,差异具有显著性意义。

辨析:从原文摘要结合表 2 的资料可以看出,该实验共涉及到 3 个因素,分别为缺血再灌注与否、时间点及软骨层。缺血再灌注与否有 2 个水平,对应于缺血再灌注组和对照组;时间点有 8 个水平;软骨层有 2 个水平,分别是浅层和中层。此处缺血再灌注与否属于实验因素,软骨层是一个重复测量因素,同一只大鼠在软骨浅层和中层中 HSP70 的表达灰度可能是相关的。再来看时间因素,如果所有的大鼠在每个时间点都进行了检测,那么时间也是一个重复测量因素,同一只大鼠不同时间点的检测数据可能不独立,此时该资料为具有两个重复测量的三因素设计定量资料;如果每只大鼠只在某一个时间点检测,则时间不属于重复测量因素,这时的设计类型是具有一个重复测量的三因素设计。总之,无论时间是否属于重复测量因素,由于软骨层这个因素的存在,该资料都应该选用重复测量定量资料的方差分析。原文中对各时点关节软骨浅层及中层 HSP70 表达灰度使用配对 t 检验进行比较,这样的做法是不合适的。此外,定量资料的正确表达方式是均数±标准差,原表中没有提供标准差这一项。

例 5:某研究^[8]的目的是评价活血止痛汤预防实验性椎板切除术后硬膜外瘢痕形成与粘连的疗效,将雄性 SD 大鼠 90 只随机分成空白对照组(A)、透明质酸钠组(B)、活血止痛汤组(C),每组 30 只,每组分别于术后 2、4、8 周末脱臼法各处死 10 只动物,测定其硬膜外瘢痕组织中羟脯氨酸(Hyp)含量,按 Rydell 法评定粘连程度。羟脯氨酸含量数据见表 3,数据的比较采用单因素方差分析和 q 检验。

辨析:本例中需要特别注意的是时间在这里并非重复测量因素,因为原文中曾提到“每组分别于术后 2、4、8 周末脱臼法各处死 10 只动物”,也就是每只动物只在一个时间点上进

表 2 两组各时点股骨头骨骺关节软骨浅层、中层 HSP70 的表达灰度

组别	3 h		6 h		12 h		24 h		48 h		5 d		2 周		4 周	
	浅层	中层	浅层	中层	浅层	中层	浅层	中层	浅层	中层	浅层	中层	浅层	中层	浅层	中层
缺血再灌注组	106.52	102.59	120.78	114.22	109.60	110.96	156.70	150.89	150.58	158.14	148.45	156.65	116.30	104.46	111.92	102.52
对照组	108.41	101.98	104.12	114.42	121.40	120.31	117.79	102.51	120.47	118.02	102.26	103.91	110.53	105.55	104.68	104.25

表 3 不同时相各组 Hyp 含量($\bar{x} \pm s, \mu\text{g}/\text{mg}$)

组别	例数	第 2 周	第 4 周	第 8 周
A 组	30	1.16±0.30	1.34±0.25	1.50±0.18
B 组	30	0.90±0.25*	1.14±0.24 [▲]	1.38±0.19
C 组	30	0.99±0.20	1.07±0.21 ^{▲▲}	1.23±0.23*

注:与同期 A 组比较,* $P=0.021$,[▲] $P=0.046$,^{▲▲} $P=0.006$,* $P=0.001$

行了测量。要看某个因素是否为重复测量因素,关键要考察是否在该因素不同水平条件下进行了两次或两次以上的重复观测,实际工作中不能一出现时间因素,就认为是重复测量设计。研究中的另一个实验因素为药物类型,如果药物类型和时间这两个因素对观测指标的影响是同等重要的,那么本例为两因素析因设计;如果两个实验因素对观测指标的影响有主次之分,则该实验为系统分组设计(也称嵌套设计)。根据设计类型的不同,可以选用相应设计定量资料的方差分析,原文中采用单因素方差分析是不合适的。

3 列联表资料的统计分析 & 错误辨析

定性资料可以整理为列联表资料,其中四格表资料的分析较为简单,这里重点说明各种 $R \times C$ 表及高维列联表的分析。在 $R \times C$ 表中,如果是双向无序的,可以使用一般检验或 Fisher 精确检验。如果是结果变量为有序变量的单向有序 $R \times C$ 表,可选用秩和检验、Ridit 分析或有序变量的 Logistic 回归分析。若是属于双向有序且属性不同的 $R \times C$ 表,则要按分析目的确定分析方法,第一种目的是只关心各组结果变量取值之间的差别是否有统计学意义,此时仍将其视为结果变量为有序变量的单向有序 $R \times C$ 表;第二种目的是研究原因变量和结果变量之间是否存在相关关系,采用 Spearman 秩相关分析;如果两个有序变量之间有相关关系,希望进一步知道这两个变量之间是否呈直线变化关系,这时需要进行线性趋势检验;如果想考察各行上的频数分布是否相同,则将资料看做双向无序的 $R \times C$ 列联表资料。当列联表是双向有序且属性相同的 $R \times C$ 表时,常用的统计分析方法为一致性检验,即 Kappa 检验。

高维列联表可以按照结果变量的取值形式分为三类,分别是结果变量为二值变量的高维列联表、结果变量为多值有序变量的高维列联表和结果变量为多值名义变量的高维列联表。对于结果变量为二值变量的高维列联表,可选用 CMH 检验、Logistic 回归分析和对数线性模型;当结果变量为多值有序变量时,可以选用 CMH 检验和有序变量的 Logistic 回归分析;当结果变量为多值名义变量时,可以选用的统计分析方法有 CMH 检验、扩展的 Logistic 回归分析和对数线性模型。

例 6:在例 5 所提到的研究^[8]中,术后硬膜外瘢痕组织与周围组织的粘连分级数据见表 4,术后第 2 周由于粘连程度较轻故未评价,资料的分析采用多个独立样本秩和检验。

辨析:本例中粘连分级为结果变量,药物类型和时间为原因变量,在使用列联表表达资料时,原因变量应该放在横标目的位置,结果变量则位于纵标目,原表中正好颠倒放置了,所以先对上表进行修改,修改结果见表 5。

该资料中粘连程度为结果变量,属于结果变量为多值有

序变量的高维列联表,可以使用 CMH 检验和有序变量的 Logistic 回归分析。仍然需要指出的是:对于时间因素而言,如果

表 4 术后第 4、8 周时各组粘连分级比较(例)

等级	4 周			8 周		
	A 组	B 组*	C 组**	A 组	B 组	C 组 [▲]
0 级	0	0	0	0	0	0
I 级	1	3	2	0	1	1
II 级	2	5	6	1	4	5
III 级	7	2	2	9	5	4

注:与同期 A 组比较,* $P=0.038$,** $P=0.048$,[▲] $P=0.021$

表 5 术后第 4、8 周时各组粘连分级比较(例)

时间	药物种类	例数				
		等级:	0 级	I 级	II 级	III 级
4 周	A 组		0	1	2	7
	B 组		0	3	5	2
	C 组		0	2	6	2
8 周	A 组		0	0	1	9
	B 组		0	1	4	5
	C 组		0	1	5	4

观察对象的粘连程度在不同时间点上进行了多次观察,则该资料是具有重复测量的定性资料;在本例中由于动物是在不同时间点被分别处死观察的,所以时间并非是重复测量因素。

考虑到受试对象首先按所接受的药物种类被分成 A、B、C 3 组,然后,每组受试对象再按药物作用时间被分成 4 周与 8 周,故在表 5 中,若能先按“药物种类”划分后再按时间划分,则更符合资料收集的实际顺序。

假如不存在时间这个因素,单就药物类型和粘连分级这两个变量来说,属于结果变量为有序变量的单向有序 3×4 表,则可以使用原文中提到的多个独立样本秩和检验来分析。

参考文献

- 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. 7.
- 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. 8.
- 蓝树华,叶方,池永龙,等. Cervifix 颈椎后路杆固定系统垂直水平悬吊改良单开门椎管成形术. 中国骨伤,2006,19(8):464-466.
- 占蓓蕾,叶舟. 保留椎体后侧壁植骨治疗多节段脊髓型颈椎病. 中国骨伤,2006,19(8):467-468.
- 胡良平. Windows SAS 6.12&8.0 实用统计分析教程. 北京:军事医学科学出版社,2001. 1.
- 张国福,王和鸣. 补阳还五汤对骨髓间质干细胞移植治疗大鼠脊髓损伤的影响. 中国骨伤,2006,19(8):452-454.
- 张敬东,陈华,温宏,等. 股骨头骨缺血再灌注关节软骨诱导型一氧化氮合成酶及热休克蛋白 70 表达的免疫组化研究. 中国骨伤,2007,20(8):537-539.
- 夏志敏,周辉,赵万军,等. 活血止痛汤对硬膜外瘢痕组织中羟脯氨酸含量及其超微结构的影响. 中国骨伤,2007,20(8):532-535.

(收稿日期:2008-07-14 本文编辑:李为农)