

《中国骨伤》杂志论文中存在的统计学错误辨析(4)

程瑞专, 胡良平, 葛毅, 李长平, 高辉
(军事医学科学院生物医学统计学咨询中心, 北京 100850)

本文针对《中国骨伤》杂志 2006 年第 4 期、2007 年第 4 期刊载文章中的统计学应用中出现的一些典型错误进行辨析, 希望能从生物医学统计学应用的角度, 对改进科研工作、提高论文质量有一定帮助。

1 实验设计与典型错误辨析

实验设计要点概括起来就是“三要素、四原则和设计类型”。“三要素”即受试对象、实验因素和实验效应。实验设计的“四原则”, 即随机、对照、重复、均衡原则。实验设计类型就是用来安排实验因素(含区组因素)及其水平的具体结构或模式。首先要了解和掌握各种实验设计类型及其特点, 才能正确选用实验设计类型, 通过经济、高效的方法实现实验目标^[1]。

1.1 实验设计时应该选用合理的实验设计类型, 从而对实验的整个过程进行质量控制 例 1 《补肾益骨方对去势大鼠血生化、骨量和骨力学性能的影响》^[2]一文, 将 40 只 Wistar 雌性大鼠, 随机分为模型组(A)、治疗组(B)、阳性对照组(C)、正常对照组(D)4 组, 每组 10 只, D 组行假手术(单纯剖腹关闭), 余 3 组行完整双侧卵巢摘除术, 术后常规饲养 12 周。A、D 组以蒸馏水 10 ml/kg 灌胃, B 组以 10 ml/kg 补肾益骨方提取液灌胃, C 组以 10 ml/kg α-D₃ 胶丸水溶液灌胃, 用药 12 周后观测某些相关指标。计量资料的比较采用单因素方差分析, 现摘录部分观测数据见表 1。

表 1 补肾益骨方提取液对去势大鼠右股骨上端骨密度变化的影响($\bar{x} \pm s, n=10$)

组别	BMC(g)	BMD(g/cm ²)
A	0.002 3±0.001 9 ^{●●}	0.006 6±0.005 8 ^{●●}
B	0.021 6±0.004 8 ^{▲▲}	0.051 3±0.012 3 ^{▲▲}
C	0.019 6±0.007 7 ^{▲▲}	0.053 3±0.020 9 ^{▲▲}
D	0.018 1±0.005 2 ^{▲▲}	0.049 4±0.015 0 ^{▲▲}

注: 与 A 组比较, ▲▲P<0.01; 与 B 组比较, ●●P<0.01

请问, 原文在实验分组及整个实验设计中存在什么问题? 应该怎样做?

辨析: 表 1“组别”下的“A、B、C、D”4 组, 容易被看作一个因素的 4 个水平组, 把本实验当作单因素 4 水平设计。但分析

实验分组情况就可以看到, 本实验涉及两个因素, 一个因素是“是否去势”, 有“是”和“否”两个水平, 另一个因素是用何种药物灌胃, 有“用蒸馏水灌胃”、“用补肾益骨方提取液灌胃”和“用 α-D₃ 胶丸水溶液灌胃”3 个水平, 希望考察这两个因素的独立作用及因素间的交互作用, 应该选用析因设计。析因设计是将实验中涉及到的实验因素各水平全面组合形成实验条件, 在各实验条件下, 做 K 次独立重复实验。一个 2 水平实验因素和一个 3 水平实验因素全面组合, 应该有 6 个实验条件, 每个条件下至少做 2 次重复实验, 即“去势与否”和“用何种药物灌胃”全面组合形成 6 个实验组: ①“去势+蒸馏水灌胃组”; ②“去势+补肾益骨方灌胃组”; ③“去势+α-D₃ 胶丸水溶液灌胃组”; ④“不去势+蒸馏水灌胃组”; ⑤“不去势+补肾益骨方灌胃组”; ⑥“不去势+α-D₃ 胶丸水溶液灌胃组”。用析因设计的标准型表格表达资料(见表 2)。

结合表 2 来看原文分组情况, 可以看到原文实验设计时少了⑤和⑥两个对照组, 即缺少表 2 中“?”位置的数据, 属于对照不全, 根本原因在于不熟悉实验设计类型, 未进行周密的科研设计, 误用单因素多水平设计代替析因设计, 由此产生的问题: ①收集资料不全面, 不能准确达到实验目的; ②分析结果不严密, 不能令人信服; ③本来仅用 6×2 个样本的实验, 就能得到关于两因素及其交互作用对结果影响的评价, 却用了 40 个样本, 还仅能得到残缺不全的信息和证据不足的结论。

1.2 随机分组不是“随意”分组, 应该使用统计学方法对受试对象的分组进行合理控制 例 2 《桡骨远端骨折的个体及其临床意义》^[3]一文, 为观察不同治疗方法对不同类型桡骨远端骨折的临床治疗效果, 将在 1999 年 6 月到 2003 年 1 月院内收治的 198 例桡骨远端骨折患者按 Cooney 桡骨远端骨折的国际分类系统分为 I-IV 型, 然后采用随机分组抽签法把每型患者分为手术治疗组和保守治疗组(如 I 型分组, 编制 50 枚号签, 充分混合后取 25 枚为手术治疗组, 剩下 25 枚为保守治疗组, 编号以入院时间的先后为依据), 各型组间年龄(t 检验)及性别(卡方检验)无统计学差异(P>0.05)。经过相应治疗和出院后随访, 观测相关指标来评价两种疗法的优劣。

请问, 原文在对患者分组时存在什么缺点? 合适的做法是

表 2 补肾益骨方提取液对去势大鼠右股骨上端骨密度变化的影响($\bar{x} \pm s, n=10$)

用何种药物灌胃	BMC(g)			BMD(g/cm ²)	
	去势与否:	去势	不去势	去势	不去势
蒸馏水		0.002 3±0.001 9	0.018 1±0.005 2	0.006 6±0.005 8	0.049 4±0.015 0
补肾益骨方灌		0.021 6±0.004 8	?	0.051 3±0.012 3	?
α-D ₃ 胶丸水溶液		0.019 6±0.007 7	?	0.053 3±0.020 9	?

什么?

辨析:原文在各型分组时,按入院顺序采取抽签法随机将各型患者分为两组,采取的是完全随机化。实现完全随机化有多种,可以查“随机数字表”、“随机排列表”和用计算机产生的“伪随机数字表”等。但完全随机化是否效果最好?不一定!随机化的作用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较好的均衡性,提高实验资料的可比性和结论的说服力。原文在随机化分组和控制重要的非实验因素方面考虑欠缺,患者个体间差异比较大,如果各类型患者数量不够大,这种完全随机的分组方法可能使一些非实验因素严重影响两组患者疗效的比较,从而影响最终结论。哪些非实验因素可能对观测结果影响较大,需要从专业上考虑,如年龄、性别、病情、患病及明确诊断时间、全身性疾病、营养状况等,原文涉及的非实验因素只有年龄和性别,考虑的不够完善。本文实验分组为达到“随机、均衡”原则,可以用“最小不平衡指数法”(详见相关专著),这是临床实验中常用的随机化分组方法,新的受试对象进入哪一组,要看进入哪一组计算出的“不平衡指数”最小,从而在分组时就对这些非实验因素进行了合理控制,达到组间比较均衡性要求^[1]。目前,在一些临床实验中,事先未对受试对象的分组进行人为控制,而是以“随意代替随机”先得到实验数据,再用统计学方法对组间的均衡性进行分析,结果只能是亡羊补牢。

从以上统计学应用错误辨析中可以看出,一个合理的实验设计方案,在进行实验之初,就应制定合理的实验方案,着重考虑实验设计“三要素、四原则”是否正确,选择合适的实验设计类型,使统计学方法的选择与应用贯穿于整个科研过程之中,良好的专业基础加统计学的思维指导和正确应用,科研质量才能得到保证。

2 统计分析方法和典型错误辨析

2.1 正确辨析实验资料性质,是选用正确统计分析方法的首要前提 例 3 《前路联合三维矫形手术治疗青少年特发性僵硬型胸椎侧凸》^[4]一文中,青少年特发性僵硬型胸椎侧凸 34 例,均采用经前路松解植骨融合联合后路三维矫形内固定治疗。术后随访,术前术后均观测 Cobb 角、顶椎旋转度、下融合椎旋转度、躯干偏移距离等指标(见表 3)。

采用 SPSS 12.0 统计软件,手术前后 Cobb 角矫正与椎体旋转度变化的比较采用 *t* 检验,数据以均数±标准差表示,Cobb 角矫正率的比较采用 χ^2 检验。

请问,表 3 中的“矫正率”适合用 χ^2 检验分析吗?

辨析:资料类型现代划分方法从资料的性质角度将资料分为“定量资料和定性资料”两大类,一般来说,定量资料是观

表 3 手术前后 Cobb 角变化和椎体旋转度变化($\bar{x}\pm s$)

项目	术前	术后	矫正率(%)
Cobb 角(°)	82.76±16.10	16.15±11.46	80.62
顶椎旋转度(°)	3.10±0.58	1.15±0.16	62.91
下融合椎旋转度(°)	1.82±0.58	0.95±0.16	47.60

察单位某项指标的具体数值,而定性资料是具有某一相同属性个体的个数^[5]。人们一提到“率”,就不加思索地认为是相对数,是定性资料,其实不然,这要看“率”是怎么得到的。原文提到表 3 中的“矫正率”=[(术前度数-术后度数)/术前度数]×100%,可见“矫正率”是从每名患者身上测得、经过计算得到具体数据,而不是计数患者个数得到的资料,应为定量资料。而且,表 3 中的 3 个矫正率应分别代表 3 个观测指标的平均矫正率,还缺少与其对应的标准差数值。 χ^2 检验是处理定性资料的统计分析方法,用定性资料的分析方法来处理定量资料显然不合适。本文对患者手术前后 Cobb 角的差异采用配对设计定量资料的 *t* 检验分析,其实相当于对 Cobb 角矫正率的分析,而再对 Cobb 角矫正率行 χ^2 检验则似有画蛇添足之嫌。由此可见,统计软件只是人们进行统计分析的工具,至于资料性质的判定和统计分析方法的合理选择还需要操作者本身掌握和理解,生搬硬套得出的结论可信度不高,有时甚至容易得出错误的结论。

2.2 正确运用统计分析方法,才能从实验数据中找到业已存在的客观规律 例 4 《关节镜下膝关节内骨折复位内固定的临床应用研究》^[6]一文中,关节镜监视下治疗膝关节内骨折 157 例,胫骨平台骨折 78 例,髌骨骨折 24 例,胫骨髁间棘骨折 36 例,股骨髁骨折 19 例。术后随访,采用 ROM 和 Lysholm 评分标准评价临床疗效,其优良率分别是 87.9%和 90.4%。结论,关节镜下微创手术为治疗膝关节内骨折提供了一种新方法,具有创伤小、并发症少、早期功能锻炼、术后功能恢复好等优点。实验所得数据见表 4。

请问,该实验结论存在什么问题?正确的做法是什么?

辨析:纵观全文,只是对实验数据进行了简单的统计描述,未采用任何统计分析方法对数据进行处理,这是缺乏统计思想的体现,仅凭数据表面现象并不能说明问题,所得结论缺乏可信性。表 4 资料属于结果变量为有序变量的单向有序 4×4 列联表资料,可选用的统计分析方法有:秩和检验或 Ridit 分析等,可以分析出采用这种疗法对各类膝关节内骨折治疗效果之间的差别是否具有统计学意义。

2.3 未弄清定量资料所对应的实验设计类型,盲目套用统计分析方法得出的结论缺乏科学性 例 5 《大鼠皮神经卡压模

表 4 157 例膝关节内骨折患者的 Lysholm 与 ROM 疗效评价(例)

Position of fracture	Case	Lysholm					ROM				
		Excellent	Good	Fair	Poor	Excellent and good rate(%)	Excellent	Good	Fair	Poor	Excellent and good rate(%)
Tibial plateau	78	41	28	7	2	88.5	39	29	7	3	87.2
Patella	24	19	4	1	0	95.3	20	2	2	0	91.7
Intercondylar spine of tibia	36	25	9	2	0	94.4	26	6	4	0	88.9
Frmoral condyle	19	10	6	2	1	84.2	9	7	1	2	84.2

表 5 手术治疗与保守治疗的 GW 评分结果

Type	Conservative group			Surgical group		
	Subjective	Objective	Total	Subjective	Objective	Total
I	0.92±0.02	1.06±0.01	2.00±0.56	1.00±0.01	1.01±0.02	2.07±0.45
II	2.31±0.12	4.25±0.08	6.56±0.18	1.16±0.02	2.02±0.05	3.18±0.18
III	2.02±0.10	5.94±0.21	8.00±0.83	2.36±0.18	5.68±0.81	8.04±1.12
IV	3.52±0.15	9.87±0.24	12.38±1.10	3.67±0.51	5.34±0.32	9.00±0.95

型的实验研究^[7]一文,借鉴 Mackinnon 设计的周围神经卡压模型所采用的方法,分别用不同内径的硅胶管套在大鼠的隐神经上,以产生不同程度的皮神经卡压。将 40 只 Wistar 雌性大鼠随机分为 5 组,每组 8 只,A 组:隐神经不做任何处理;B 组:游离隐神经,不进行卡压处理;C 组:以内径 0.3 mm 的硅胶管卡压神经;D 组:以内径 0.4 mm 硅胶管卡压神经;E 组:以内径 0.5 mm 硅胶管卡压神经。术后 1 周观测相关指标,其中,电生理检测观测指标为神经传导速度,结果,A、B、E、D 各组神经传导速度分别为(43.2±1.8)、(40.4±2.4)、(33.6±2.7)、(22.3±3.1) m/s,C 组均未能引出动作电位。两样本均数之间的比较用 *t* 检验,神经传导速度 B 组小于 A 组($P<0.01$),E 组和 D 组均小于 A 组($P<0.01$),D 组小于 E 组($P<0.01$),显示出随着硅胶管内径的缩小,压力的增加,神经传导速度逐渐减慢的趋势。

请问,原文在统计方法上存在什么问题?

辨析:从实验设计类型来看,原文不属于任何标准设计类型,可称之为多因素非平衡的组合实验,应进行合理拆分,找出在专业上和统计学上都成立的实验组合。拆分结果:组合一,A 组和 B 组形成单因素成组设计,实验因素是“是否游离隐神经”,可以用成组设计定量资料 *t* 检验进行处理;组合二,B、C、D、E 组形成单因素 4 水平设计,实验因素是“卡压隐神经的硅胶管内径”,有“0、0.3、0.4、0.5 mm”4 个水平。*t* 检验不适合分析组合二这种设计类型下收集的定量资料,*t* 检验仅能分析单因素且水平数≤2 的定量资料。对于组合二所得定量资料,如果符合方差分析的前提条件(独立性、正态性和方差齐性),可以用单因素多水平设计定量资料方差分析进行处理,如果分析结果显示各组观测指标的均数间差异具有统计学意义,可以进一步采用 LSD 法、Bonferroni 法、SNK 法(*q* 检验)等对任何两组间均数行两两比较(注意:所采用的两两比较方法不同,控制的误差类型不同)。值得注意的是:任何两组间均数的两两比较,并不能回答“随着硅胶管内径的缩小,压力的增加,神经传导速度呈逐渐减慢的趋势”这个问题;如果确实需要寻找“硅胶管内径及压力”和“传导速度”的关系,可能需要采用简单直线相关与回归方法等方法。

例 2^[3]一文中,术后随访,根据 GW 评分标准,分为主、客观两部分给予评分,结果见表 5。

用 *t* 检验处理表 5 资料,结果显示,对于 I、III 型关节内、外无移位骨折手术的组的 GW 评分要相对高一些,但两种疗法之间无统计学差异,对于 II、IV 型关节内、外位移骨折保守治疗且 GW 评分要高于手术组,两者间差异有显著性意义。

请问,原文处理表 5 资料的统计分析方法正确吗?

辨析:表 5 资料涉及 3 个因素,分别是骨折类型、治疗方

法和评分方式,其中评分方式一项由主观和客观对同一患者分别进行评分,属于重复测量因素,因此表 5 资料属于具有一个重复测量的 3 因素定量设计定量资料。*t* 检验只适用于处理满足参数检验前提条件且为单因素 $K(K\leq 2)$ 水平设计的定量资料,用在此处不合适。表 5 资料如果满足方差分析的前提条件,可采用具有一个重复测量的 3 因素设计定量资料方差分析进行处理,可以分析出各个因素及因素间可能存在的交互作用的效应大小。值得一提的是,表 5 编制得不够规范,表的左边应列出两列,分别代表骨折类型和治疗方法两个实验分组因素,表头横线上方应写“GW 评分结果($\bar{x}\pm s$)”。

例 6《踝关节生物力学实验放置压敏片入路新方法》^[8]一文,采用 6 具防腐尸体标本,左足组为 A 组,从内踝放入压敏片,右足组为 B 组,从外踝放入压敏片,分别从踝关节前内、前外、后内、后外分析压敏片,观测关节接触面积、平均压强、峰值压强等指标,部分数据见表 6。

表 6 两组踝关节平均压强比较($\bar{x}\pm s$, Mpa)

Group	Intraanterior	Extraanterior	Intralateral	Extralateral
A	2.09±0.46	2.30±0.48	1.73±0.41	1.87±0.43
B	2.11±0.47*	2.33±0.45*	1.70±0.38*	1.82±0.40*

注 * $P>0.05$

用 SPSS 10.0 软件进行方差齐性检验,LSD 法多种比较 $P>0.05$,结果,A、B 组实验踝关节平均压强差异无统计学意义。

请问,原文分析资料所采用的统计分析方法合适吗?

辨析:原作者将左足与内踝、右足与外踝视为两个分组条件,实际上漏掉了“左足与外踝、右足与内踝”这样的两个分组条件。原文实际上涉及的因素有 3 个,即“左足还是右足”、“内踝还是外踝”、“每个踝关节的不同部位”,每个因素的水平数相乘 $2\times 2\times 4=16$,应该有 16 个实验条件,每具防腐尸体标本都应在这 16 个条件下被重复观测 3 项定量指标的数值,这个设计类型应叫做具有 3 个重复测量的三因素设计;即使按原作者的想法来设计,列表成表 6 的形式,它应叫做具有一个重复测量的两因素设计。原文所采用的统计分析方法是不妥的。实验设计类型判定之后,应根据定量资料是否满足参数检验的前提条件,选用与设计类型对应的统计分析方法处理定量资料为宜。

2.4 研究设计类型直接反映实验目的,最终影响结果的解释和结论的陈述 例 7《颈椎旋转、半失稳在椎动脉型、交感型、神经根型颈椎病的相关研究》^[9]一文,颈椎病患者 112 例,分为椎动脉型 38 例、交感型 36 例、神经根型 38 例,测量患者每个颈椎体的旋转度和椎体半失稳的位移距离。分别用方差分析和卡方检验处理相关资料。结果,在椎体半失稳的位移距离

和椎体半失稳率上,椎动脉型和交感型颈椎病与神经根型颈椎病有统计学差异。结论,在椎动脉型颈椎病和交感型颈椎病中椎体半失稳及颈椎旋转是它们的发病的一个重要因素,而在神经根型颈椎病中不是发病的重要因素。原文用文字形式表达实验资料,此处将部分数据转换为表格形式(见表 7)。

表 7 各型颈椎病椎体失稳情况

颈椎病类型	患者人数	
	椎体情况:	
椎动脉型	失稳	未失稳
交感型	28	10
神经型	24	12
	10	28

请问,原文根据统计分析结果得出的结论存在什么问题?

辨析:统计研究设计类型通常分为“观察性研究”和“实验性研究”,观察性研究设计又可以具体分为:横断面研究设计、队列研究设计和病例对照研究设计,原文设计属于横断面研究设计。采用这种设计,经过统计分析,由“各类型颈椎病之间椎体失稳率及椎体旋转的差异均具有统计学意义、椎动脉型和交感型颈椎病的椎体失稳率及椎体旋转均比神经根型颈椎病高”的分析结果,无法得出“椎体失稳及椎体旋转是椎动脉型和交感型颈椎病发病因素,不是神经根型颈椎病发病因素”的推论。横断面研究设计,只是了解受试对象一些指标在某一时间所处的水平,往往不能准确判定某些事件发生的时间先后关系,也就是说,在原文中无法准确判定是先有颈椎病还是先有椎体失稳及椎体旋转,无法确定两者间的因果关系。但是这种研究设计可以大致了解事物目前现状,为进一步病因的研究提供线索^[1]。如果想为病因提供有力证据,应采用队列研

究设计,在已知椎体失稳及椎体旋转的情况下,研究者通过长时间追踪观察大量此类对象,了解是否发生颈椎病和发生何种类型颈椎病等情况,这样得出的病因结论才更具有科学性和严谨性。

目前,医学论文中统计学的应用错误比较普遍,在一定程度上影响了医学科研的质量和水平,科研人员需要对统计学的基本原理和方法更深入学习和掌握,在实际工作中进一步提高正确运用统计学的能力。

参考文献

- 1 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,2007.11-43.
- 2 杨林,姚新苗,黄竞,等. 补肾益骨方对去势大鼠血生化、骨量和骨力学性能的影响. 中国骨伤,2007,20(4):224-226.
- 3 闫合德,高伟阳,李志杰,等. 桡骨远端骨折的个体及其临床意义. 中国骨伤,2006,19(4):211-213.
- 4 闫伟强,贺西京. 前后路联合三维矫形手术治疗青少年特发性僵硬型胸椎侧凸. 中国骨伤,2007,20(4):247-250.
- 5 胡良平,李子建. 医学统计学基础与典型错误辨析. 北京:军事医学科学出版社,2003. 15-42,242-343.
- 6 陈述祥,康乐,刘红光,等. 关节镜下膝关节内骨折复位内固定的临床应用研究. 中国骨伤,2006,19(4):203-206.
- 7 张苇男,仇树林,董福慧. 大鼠皮神经卡压模型的实验研究. 中国骨伤,2007,20(4):217-219.
- 8 韩张伶,姚双权,刘勇,等. 踝关节生物力学实验放置压敏片入路新方法. 中国骨伤,2006,19(4):223-225.
- 9 熊炎炆,杜宇,陈伟珍. 颈椎旋转、半失稳在椎动脉型、交感型、神经根型颈椎病的相关研究. 中国骨伤,2006,19(4):207-209.

(收稿日期:2008-04-07 本文编辑:李为农)

中国康复医学会颈椎病专业委员会第十次学术年会
征文及会议通知

由中国康复医学会主办,山东省康复医学会、山东大学齐鲁医院承办的“颈椎病专业委员会第十次学术年会议”将于 2008 年 10 月 24 日至 27 日在济南召开。会议同时举办医学继续教育课程“腰椎间盘突出症的非手术治疗进展学习班”,授予 I 类继续教育学分 10 分。

会议内容:专题报告、大会交流和技术应用讲座。主要对近年来颈椎病领域的临床、基础科研成果及经验进行总结。

征文内容:颈椎病手术治疗新技术、新进展;颈椎病非手术治疗及康复;颈椎病中西医临床治疗及研究成果;颈椎病的基础研究、应用研究;其他脊柱疾病及脊柱相关疾病基础研究及康复治疗进展。

征文要求:欢迎 E-mail 投稿,E-mail:shouweiy@sina.com 1000 字结构式摘要,编排顺序为:目的、方法、结果、结论。注明作者姓名、单位及邮编。截稿日期为 2008 年 9 月 20 日。来稿请务必附通讯地址、联系电话、手机、E-mail,以便及时进行联系。

联系地址:济南市文化西路 107 号山东大学齐鲁医院康复科,250012。

联系人:殷翠萍,岳寿伟;联系电话:0531-82169362。