

· 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(12)

高辉, 胡良平, 郭晋, 李长平

(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, 2009, 22(1):72-75 www.zggszz.com

在生物医学领域, 统计学的重要性得到了越来越多人的认可。科研工作者自觉地将专业与统计学相结合来制定合适的实验设计方案, 并使用统计学方法表达和分析实验数据, 从而得出更有科学性、更具说服力的结论。但由于实际问题的复杂性及科研工作者的主观原因, 也导致了统计学误用的屡屡发生, 这在统计分析方面尤其突出。本文以《中国骨伤》杂志 2006 年与 2007 年第 12 期发表的部分论文数据为基础, 给出若干种科研工作者常用的错误做法, 并分析这些方法的不妥之处, 最后给出适合此资料的正确的统计分析方法。

1 资料的分类与分型

统计分析的对象是统计资料, 而正确识别统计资料的类型是合理选用统计分析方法的首要前提。资料类型的现代划分方法^[1], 是先将资料分为定量资料与定性资料两类。其中, 测量每个观察单位某项指标量的大小所得到的资料称为定量资料; 而观察每个观察单位某项指标质的状况所得到的资料称为定性资料。定量资料又可细分为计量资料与计数资料, 指标的取值可以带度量衡单位甚至可以带小数的定量资料称为计量资料; 指标的取值可以带度量衡单位但不可以带小数的定量资料称为计数资料。定性资料又可细分为名义资料与有序资料, 指标质的不同状况之间在本质上无数量大小或质量好坏之分或先后顺序之分的定性资料称为名义资料; 指标质的不同状况之间(状态个数≥3)在本质上无数量大小或质量好坏或先后顺序之分的定性资料称为有序资料。

2 正确处理定量资料

分析定量资料, 关键点在于结合研究者的研究目的, 正确判定资料对应的实验设计类型, 并根据资料是否满足参数检验的前提条件, 选用正确的统计分析方法。切忌不考虑资料的实际情况, 盲目套用 *t* 检验处理一切定量资料。

例 1, 《陆文整骨三步五法治疗腰椎间盘突出症临床规范化研究》^[2]—文中, 研究者将 233 例腰椎间盘突出症患者随机分为两组: 陆文整骨手法组(研究组)122 例, 热敷牵引组(对照组)111 例。分别记录患者手术前和手术后疼痛、腰部活动度(LVM)、直腿抬高高度(SLR)及症状、体征和日常生活活动

能力情况。其中, 疼痛状况采用疼痛视觉模拟评分法(VAS)进行评估, 症状、体征和日常生活活动能力状况采用改良日本骨科学会下腰痛功能评价表(M-JOA)进行评估。实验结果见表 1, 若对两组患者在治疗前及治疗后 VAS、LVM、SLR、M-JOA 的差异分别采用成组设计定量资料的 *t* 检验进行分析, 是否正确?

释疑: 研究者欲研究陆文整骨三步五法与热敷牵引疗法在治疗腰椎间盘突出症的体征改善方面是否存在差异。实验因素为“是否采用陆文整骨手法”或“治疗方法”, 受试对象为腰椎间盘突出症患者, 观测指标为疼痛、腰部活动度的评分, 症状、体征和日常生活活动能力的评分及直腿抬高高度。本研究中, 研究者在治疗前分别测量了两组患者的有关指标情况, 为平衡两组患者在治疗前测量指标取值上的差异, 可以治疗前这些指标的数值作为“基础值”, 采用相应设计类型定量资料的协方差分析。233 例腰椎间盘突出症患者被随机分为两组, 分别接受两种不同的方式治疗, 其实验设计类型应为成组设计定量资料或称单因素两水平设计定量资料, 考虑到治疗前观测指标取值这个定量的影响因素的存在, 对此资料选用成组设计定量资料的协方差分析为宜。若分别对两组患者在治疗前、治疗后每个观测指标的取值采用成组设计定量资料的 *t* 检验进行分析, 认为“若两组患者在治疗前某指标的取值上无统计学差异, 而在治疗后两组患者此指标的取值上有了统计学差异, 所以可判定两种治疗方法的疗效之间的差异有统计学意义”, 则是没有注意到在治疗前和治疗后测自同一患者的某指标的两个观测值之间存在内在的联系, 可能会得出错误的结论。当然, 若资料不满足参数检验的前提条件, 则应进行适当的变量转换, 并对转换后满足参数检验前提条件的数据进行分析。

例 2, 《干扰电疗法治疗腰椎间盘突出症疗效的对照研究》^[3]—文中, 研究者欲评价牵引加干扰电疗法治疗腰椎间盘突出症的临床效果, 并与单纯牵引疗法相比较。选取 180 例腰椎间盘突出症患者, 考虑到病程的影响, 采用分层随机分组的方法, 将 180 例患者分入牵引加干扰电疗法治疗组(B 组)和

表 1 治疗前后临床体征改善情况

组别	例数	手术前				手术后			
		VAS	LVM	SLR	M-JOA	VAS	LVM	SLR	M-JOA
研究组	122	8.3±1.2	133.7±56.3	42.7±16.1	20.7±5.7	2.4±1.9	211.6±39.3	68.1±9.6	7.8±4.7
对照组	111	7.7±2.0	134.2±55.0	44.8±18.7	19.4±6.4	3.1±2.3	196.2±48.3	66.8±11.0	9.2±5.3

表 2 两组不同病程治疗效果($\bar{x}\pm s$,分)

组别	病例数	<2 个月			2 个月~2 年			>2 年		
		治疗前	治疗 7 d	治疗 14 d	治疗前	治疗 7 d	治疗 14 d	治疗前	治疗 7 d	治疗 14 d
A 组	89	13.6±2.44	23.8±2.45	28.3±1.57	13.4±2.41	21.8±2.25	26.4±1.55	12.8±2.61	20.1±2.33	25.4±1.44
B 组	91	13.2±2.37	26.1±2.19	29.4±1.42	13.6±2.27	24.0±2.16	27.7±1.62	12.3±2.34	21.2±2.06	26.2±1.71

表 3 两种疗法治疗不同病程腰椎间盘突出症患者的效果

治疗方法	病程	病例数	疗效($\bar{x}\pm s$,分)		
			观察时间: 治疗前	治疗 7 d	治疗 14 d
单纯牵引疗法	<2 个月	n_1	13.60±2.44	23.80±2.45	28.30±1.57
	2 个月~2 年	n_2	13.40±2.41	21.80±2.25	26.40±1.55
	>2 年	n_3	12.80±2.61	20.10±2.33	25.40±1.44
牵引加干扰电疗法	<2 个月	n_4	13.20±2.37	26.10±2.19	29.40±1.42
	2 个月~2 年	n_5	13.60±2.27	24.00±2.16	27.70±1.62
	>2 年	n_6	12.30±2.34	21.20±2.06	26.20±1.71

单纯牵引治疗组(A组)。对治疗效果采用日本矫形外科学会制订的“腰椎疾患治疗成绩评分表”进行评分,分别记录患者在治疗前、治疗 7 d 及 14 d 后的疗效评分,结果见表 2。若对各病程、各时点上的数据分别采用成组设计定量资料的 t 检验进行分析,妥否?

释疑:研究者欲评价牵引加干扰电疗法与单纯牵引疗法治疗腰椎间盘突出症的临床效果,分组时考虑到病程的影响,采用分层随机分组的方法,并记录每位患者在治疗前、治疗 7 d 及 14 d 3 个不同的时点上的疗效评分。本研究中,涉及到的因素有 3 个,分别为治疗方法、患者病程及观测时间。其中,同一患者在 3 个不同的时点上分别被观测其疗效评分,因此观测时间应为重复测量因素。所以,此资料对应的设计类型应为具有一个重复测量的三因素设计,可选用具有一个重复测量的三因素设计定量资料方差分析。当然,若以治疗前疗效评分为“基础值”,采用具有一个重复测量的三因素设计定量资料的协方差分析更佳。

此外,表 2 在表达资料时存在一些不足。观测指标在表中没有直接地体现出来,且病程占据了总纵标目的位置,标目设置不当。A 组的 89 例患者应为此组中 3 个病程患者例数之合计,B 组亦是如此,但表 2 目前的表述会使读者误以为 A 组每个病程的患者都为 89 例,B 组每个病程的患者也均为 91 例。本资料表达的“标准型”见表 3,其中 n_1-n_6 分别代表 A、B 两组不同病程的患者例数,由于从表 2 中无法获取其具体数值,故以符号代替。

3 正确处理定性资料

分析定性资料,关键点在于注意结合研究目的,正确识别列联表的类型,并严格遵守统计分析方法的应用条件。如结合列联表类型和研究目的认为可选用 χ^2 检验分析资料时,还要检查其是否满足一般 χ^2 检验的前提条件;若列联表中有超过 1/5 的格子的理论频数小于 5 时,应选用校正的 χ^2 检验或 Fisher 精确检验。

例 3,资料同例 1。研究者同时采用改良日本骨科学会下腰痛功能评价表(M-JOA)对疗效进行比较分析。以改善率表示,定义改善率=[(治疗前分值-治疗后分值)/治疗前分值]×

100%,改善率 $\geq 75\%$ 为优, $\geq 50\%$ 、 $<75\%$ 为良, $\geq 30\%$ 、 $<50\%$ 为可, $<30\%$ 为差。其中,优、良、可评定为有效,差评定为无效,结果见表 4。若欲比较两组患者有效率之间的差异是否有统计学意义,应如何分析?若欲比较两组患者整体疗效之间的差异是否有统计学意义,应如何分析?

表 4 临床疗效的 M-JOA 评价

组别	病例数	优	良	可	差	有效率
研究组	122	36	53	19	14	88.5%
对照组	111	22	34	27	28	74.8%

释疑:若比较两组患者有效率之间有无统计学差异,则需将 M-JOA 疗效评价为优、良、可的患者例数合并,与疗效评价为差的患者例数一起重新列表,见表 5,此资料应为队列研究设计的 2×2 表资料。欲比较两个治疗组有效率之间的差异是否有统计学意义,首先应将此资料视为横断面研究设计的 2×2 表资料进行处理,因为表 5 中任一格子的理论频数均大于 5,故可采用 χ^2 检验比较两组患者有效率之间有无统计学差异;若得到两组患者有效率之间的差异有统计学意义时,再计算相对危险度 RR 值,然后采用 $MH\chi^2$ 检验来分析总体 RR 值与 1 之间的差异是否有统计学意义。另外,若比较两组患者整体疗效之间有无统计学差异,则此资料应视为结果变量为有序变量的 2×4 列联表资料,可采用秩和检验分析此资料。

表 5 临床疗效的 M-JOA 评价

治疗方法	例数(人)	
	疗效: 有效	无效
陆文整骨手法	108	14
热敷牵引疗法	83	28

例 4,《双侧臀肌挛缩症的手术治疗》^[4]一文,研究者欲探讨经股骨大转子弧形向上后方延长至条索状切口手术治疗臀肌挛缩症的疗效,选取 78 例患者进行手术并随访 6~36 个月观察其疗效。其中,优 58 例,良 14 例,差 4 例,优良率为 95%。具体为:年龄不超过 10 岁的 9 例患儿疗效均为优;11~15 岁患儿疗效优 39 例,良 6 例;患者中有 8 例病情较重,年

龄均超过 15 岁,4 例疗效差,4 例疗效良。请问:若不考虑病情在 3 个年龄组(<10 岁, 11~15 岁, >15 岁)中可能存在的失衡,应如何分析此资料?

释疑:根据研究者原文的表述,将资料以统计表形式给出,见表 6。可以看出,此资料中原因变量为“不同年龄组”,结果变量为“疗效”,二者均为有序变量但本质属性不同,所以此资料为双向有序且属性不同的二维列联表资料。分析此资料需结合分析目的和列联表类型来选择统计分析方法。若希望考察各组的频数分布是否相同,应将此资料视为双向无序的二维列联表资料;因表中有超过 1/5 的格子的理论频数小于 5,应选用 Fisher 精确检验。若希望考察各组间整体疗效之间的差异是否有统计学意义,应将此资料视为结果变量为有序变量的单向有序列联表资料,可选用秩和检验或 Ridit 分析。若希望考察两个有序变量之间的相关关系有无统计学意义,可采用 Spearman 秩相关分析。若希望进一步考察两变量之间的变化关系是否为直线关系时,可选用线性趋势检验。

表 6 不同年龄组臀肌挛缩症患者手术疗效结果

年龄(岁)	例数(人)			
	疗效:	优	良	差
<10		9	0	0
11~15		39	6	0
>15		10	8	4
合计		58	14	4

例 5,《三种固定方法治疗儿童肱骨髁上骨折的临床对照试验》^[5]一文,研究者欲探讨不同固定方法治疗儿童移位肱骨髁上骨折临床疗效及临床适应证,选用 137 例移位肱骨髁上骨折儿童,分别采用交叉克氏针内固定(A 组,38 例),小夹板外固定(B 组,67 例),石膏托外固定(C 组,32 例)。观察 3 组肘内翻的发生率,结果见表 7。请问:可否采用 χ^2 检验对 3 组患者肘内翻的发生率进行两两比较?

表 7 3 组患者肘内翻的发生情况

组别	病例数(例)	肘内翻	
		例数	发生率(%)
克氏针内固定组	38	2	5.26
夹板固定组	67	6	8.96
石膏固定组	32	10	31.25

释疑:表 7 中总纵标目的设置不妥,对其进行修改,见表 8。本资料原因变量为“固定方式”,结果变量为“肘内翻否”,前者为多值名义变量,后者为二值变量,资料类型为结果变量为二值变量的二维列联表资料。欲考察 3 组患者肘内翻发生率之间的差异有无统计学意义,正确的做法是选用 χ^2 检验(本资料满足 χ^2 检验应用的前提条件)来分析此 3x2 列联表资料,若得到 3 组患者肘内翻发生率之间的差异确实有统计学意义,可采用独立分割法^[6]进行两两比较,也可采用 Bonferroni 法或 Sidak 法等调整两两比较的显著性水准。若直接采用 χ^2 检验进行两两比较,就像用 t 检验进行单因素多水平设计定量资料的两两比较一样,增大了犯假阳性错误的概率。

例 6,《肱骨髁上骨折手法复位穿针和夹板固定的疗效比

表 8 3 组患者肘内翻的发生情况

固定方式	例数(人)		
	肘内翻否:	是	否
克氏针内固定		2	36
夹板固定		6	61
石膏固定		10	22

较》^[7]一文,研究者将 240 例患者随机等分为 2 组:夹板组和穿针组。其中夹板组 Gartland II 型 65 例, Gartland III 型 55 例,穿针组 Gartland II 型 65 例, Gartland III 型 55 例。两组患者治疗结果见表 9。请问:分别对 Gartland II 型和 Gartland III 型的两组患者采用 χ^2 检验分析两种治疗方式的疗效之间的差异是否有统计学意义,正确吗?

表 9 伸直型肱骨髁上骨折夹板组和穿针组的疗效

分型	治疗方法	病例数	优	良	差	优良率
Gartland II	夹板组	65	47	9	9	86.2%
	穿针组	65	52	7	6	90.8%
Gartland III	夹板组	55	48	4	3	94.5%
	穿针组	55	45	6	4	92.7%

释疑:本资料原因变量有两个,分别为“疾病分型”和“治疗方法”,均为二值变量,结果变量为“疗效”,是一多值有序变量。因此,本资料应为结果变量为多值有序变量的三维列联表资料。若分别对 Gartland II 型和 Gartland III 型的两组患者采用 χ^2 检验比较两种治疗方式疗效之间的差异是否有统计学意义,实际上是将原本为一个整体的三维列联表资料切割成了两个二维列联表资料,然后采用二维列联表资料对应的统计分析方法处理数据,这是用单因素分析法来处理同时受多因素影响的数据,割裂了多因素之间的相互联系,易得出错误的结论。此外,切割后资料就转变为两个结果变量为有序变量的二维列联表资料,分析两种治疗方法疗效之间的差异是否有统计学意义,也应选用秩和检验或 Ridit 分析而非 χ^2 检验。所以,这样随意拆分资料进行单独分析的做法是不可取的。正确的做法,应采用 CMH 校正的秩和检验或有序变量的多重 logistic 回归分析来处理此资料。

例 7,《奇正青鹏膏治疗急性痛风性关节炎的临床观察》^[8]一文,研究者选择符合 1977 年美国风湿病学会急性痛风性关节炎诊断标准,受累关节为第 1 跖趾关节的患者 72 例,随机等分为两组:试验组局部外用奇正青鹏膏同时口服非甾体抗炎镇痛药奥贝-双氯酚酸钠;对照组仅口服奥贝-双氯酚酸钠。分别于治疗后 1、3、7 d 对每例患者的临床疗效进行综合评价,结果见表 10。若对两组患者在治疗后 1、3、7 d 3 个时点上临床疗效的数据分别采用 Wilcoxon 秩和检验,以判断两组患者在这 3 个时点上临床疗效之间的差异是否有统计学意义,是否合适?

释疑:本资料较为特殊,它包含了一个时间因素,研究者是在治疗后 1、3、7 d 3 个时点上分别对每位患者的临床疗效进行综合评价,因而这里的时间因素是一个重复测量因素,所以此资料应为具有重复测量的定性资料。若对两组患者在治疗后 1、3、7 d 的临床疗效分别采用 Wilcoxon 秩和检验,则割裂

表 10 两组患者临床综合疗效评价情况比较

组别	例数	治疗后 1 d				治疗后 3 d				治疗后 7 d			
		无效	进步	有效	显效	无效	进步	有效	显效	无效	进步	有效	显效
试验组	36	9	25	2	0	0	12	10	14	0	0	15	21
对照组	36	17	19	0	0	0	27	6	3	0	4	18	14

了设计的整体性,可能会得出错误的结论,合适的做法是采用加权最小二乘法进行分析。另外,表 10 是资料表达的“表现型和原型”,没有展示出该资料重复测量设计的本质,表 11 是该资料的“标准型”。当然治疗前后的某种疗效组合可能并不存在,则相应例数为 0。如试验组和对照组治疗前后疗效分别为 1、1、1 的这种组合本资料并不存在,则其例数为 0。

表 11 两组患者临床综合疗效评价情况

奇正青鹏膏 用否	疗效			例数	
	时间:	治疗后 1 d	治疗后 3 d		治疗后 7 d
用		1	1	1	n_1
用		1	1	2	n_2
用		:	:	:	:
用		4	4	4	n_{64}
否		1	1	1	n_{65}
否		1	1	2	n_{66}
否		:	:	:	:
否		4	4	4	n_{128}

注:表内“疗效”之下 1、2、3、4 分别代表无效、进步、有效、显效

4 其他注意事项

为正确分析和有效利用实验资料的相关信息,除前述有关注意事项外,还需要在以下几个方面下足功夫。

第一,要有合适的实验结果记录表格。以重复测量设计定量资料为例,要事先编制好收集实验数据的表格,确保重复测自同一个受试者的多个数据按相应的顺序写在同一行上。否则,误将重复测自同一个受试对象的多个数据与测自其他受试对象的数据混杂在一起,就会导致误将重复测量设计当作具有独立重复实验的析因设计,从而误用统计模型,这可能会得出错误的结论。

第二,要能正确地表达资料的本质。资料的表达千变万化,不同的人有不同的表达习惯,但很多人随心所欲的表达方式不利于他人的阅读和理解。更有甚者,因为资料表达上的随

意,导致研究者自身都无法正确地识别资料的设计类型,从而采用了错误的统计分析方法处理实验数据,得出错误的结果和结论。所以,为避免这种可能“误人误己”的结果,笔者建议以“统计学三型理论^[9-10]”为指导,以“标准型”来表达获得的实验资料。

第三,要弄清因素之间的关联与关系。在多因素设计中,实验因素之间的关系必须很清晰,否则会影响到统计分析方法的正确选择。对于最终的实验条件为各因素各水平全面组合而成的实验设计,要特别注意实验因素之间是否有主次之分、实验因素进入实验过程是否存在先后之分。要根据因素及水平的组合情况、因素作用于受试对象的时间和因素对观测结果的影响重要程度等综合加以考虑,才能正确辨析出实验设计的类型,才可能对资料进行正确的分析处理。

参考文献

- [1] 胡良平. 统计学三型理论在统计表达与描述中的应用. 北京:人民军医出版社,2008. 40-42.
- [2] 陈建华,孙波,吴云定,等. 陆文整骨三步五法治疗腰椎间盘突出症临床规范化研究. 中国骨伤,2006,19(12):705-707.
- [3] 何杰民,许杰,唐勇,等. 干扰电疗法治疗腰椎间盘突出症疗效的对照研究. 中国骨伤,2006,19(12):708-710.
- [4] 徐泽孔. 双侧臀肌挛缩症的手术治疗. 中国骨伤,2006,19(12):715-716.
- [5] 方汉民,马少云,曹建斌,等. 三种固定方法治疗儿童肱骨髁上骨折的临床对照试验. 中国骨伤,2007,20(12):823-825.
- [6] 郭祖超. 医学统计学. 北京:人民军医出版社,1999. 56-59.
- [7] 潘志雄,朱永展,张兆华,等. 肱骨髁上骨折手法复位穿针和夹板固定的疗效比较. 中国骨伤,2006,19(12):719-721.
- [8] 王吉波,谢荣爱,姜秀波,等. 奇正青鹏膏治疗急性痛风性关节炎的临床观察. 中国骨伤,2006,19(12):755-756.
- [9] 胡良平,刘惠刚. 统计学的三型理论及其在生物医学科研中的应用. 中华医学杂志,2005,85(27):1936-1940.
- [10] 胡良平,刘惠刚. 统计学思想与三型理论在生物医学科研中的应用. 中西医结合学报,2007,5(2):216-219.

(收稿日期:2008-11-06 本文编辑:李为农)

作者须知

凡投稿本刊作者,请勿随信夹寄现金(包括审稿费、版面费以及任何费用)以免造成不必要误会,谢谢合作。

《中国骨伤》杂志社