

# 医学期刊论文中的析因设计、生存分析及 Bland-Altman 法常见错误分析

杨建鑫 鲁立 左琦 缪琴 闻浩 王慧琳 刘玉巧 朱琴

南京军区南京总医院《医学研究生学报》编辑部,210002,南京

**摘要** 虽然医学期刊论文中介绍统计学方法运用错误的研究已较多,如  $t$  检验、方差分析、秩和检验、 $\chi^2$  检验等,但是针对近几年来临床研究论著中不断出现的析因设计、生存分析及 Bland-Altman 法(BA 法)介绍却鲜见,上述方法的错误应用等问题容易被医学编辑人员及科研人员所忽视。该研究通过收集《医学研究生学报》中的文章,举例分析上述方法中常出现的错误以及对于如何正确运用这些方法提出建议。

**关键词** 医学论文;统计方法;常见错误

**Analysis of common statistical errors in factorial design, survival analysis and Bland-Altman of medical journal articles**

//YANG Jianxin, LU Li, ZUO Qi, MIAO Qin, WEN Hao, WANG Huilin, LIU Yuqiao, ZHU Qin

**Abstract** Although many researches have been conducted on the misuse of statistical methods, such as  $t$  testing, variance analysis, ranksum test, and  $\chi^2$  testing, in the medical journals, very few focus on the application of methods of factorial design, survival analysis and Bland-Altman. Therefore, the misuse of above three methods is very likely to be ignored by the medical editors and researchers. This paper analyses the common statistical errors of these methods with examples selected from original papers published on Journal of Medical Postgraduates and discusses how to use these methods correctly, in order to help the editors to improve the academic level of the medical journals.

**Keywords** medical thesis; statistical method; common statistical error

**Authors' address** JMP Editorial Board of Nanjing General Hospital, 210002, Nanjing, China

**DOI:**10.16811/j.cnki.1001-4314.2016.03.009

国内外研究表明,医学论文统计学方法的误用比

例相当高<sup>[1]</sup>,而是否正确运用统计学分析是判断医学论文是否具有学术价值和医学期刊质量好坏的重要标准之一<sup>[2]</sup>。虽然医学期刊论文中介绍统计学方法的研究比较多,但是近临床研究文章中常运用的方法,如析因设计、生存分析及 Bland-Altman 法(BA 法)的介绍却较少。本研究通过收集《医学研究生学报》稿件中的 3 种方法并举例分析,为医学编辑人员更好地为投稿作者提供修稿意见或退稿意见提供参考依据和建设性意见,以提高医学期刊的质量和学术水平。

## 1 析因设计

析因设计是用来分析 2 个及 2 个以上因素的各个水平的全面组合的设计,包括主效应、单独效应及交互作用<sup>[3]</sup>。

**1.1 常见错误** 在医学期刊论著类文章中常见以单因素方差分析或  $t$  检验来代替析因设计方差分析,而得出错误的结论。两者的区别主要在于前者忽略了因素之间的交互作用。当因素之间存在交互作用,则不能分析主效应,应分析单独效应<sup>[4-5]</sup>。这是投稿作者常犯的错误。

**1.2 实例分析** 在《IL-23/IL-17 轴在大鼠急性脊髓损伤模型中的表达及意义》基础研究文章中,检测 mRNA<sub>p19</sub> 在大鼠急性脊髓损伤模型中的表达水平,把雄性成年 SD 大鼠 40 只随机分为对照组和脊髓损伤组,2 组又各自分为 1、5、9 和 13 h 4 个亚组( $n = 10$  只)。原作者采用单因素方差分析(ANOVA)比较各分组之间的数据均数,并做多组间两两均数比较(LSD 检验法),见表 1。

表 1 大鼠在不同时间点 mRNA<sub>p19</sub> 的表达水平( $\bar{x} \pm s$ )

检测项目	对照组	实验组			
		1 h	5 h	9 h	13 h
mRNA <sub>p19</sub>	0.35 ± 0.02 <sup>*</sup>	0.16 ± 0.03 <sup>△</sup>	0.47 ± 0.06 <sup>●</sup>	0.31 ± 0.03 <sup>▲</sup>	0.42 ± 0.05 <sup>■</sup>

注: \* 与实验组 1、5、9、13 h 比较,  $P < 0.05$ ;  $\Delta$  与对照组、实验组 5、9、13 h 比较,  $P < 0.05$ ;  $\bullet$  与对照组、实验组 1、9、13 h 比较,  $P < 0.05$ ;  $\blacktriangle$  与对照组、实验组 1、5、13 h 比较,  $P < 0.05$ ;  $\blacksquare$  与对照组、实验组 1、5、9 h 比较,  $P < 0.05$ 。

分析:上述资料作者采用单因素方差分析只分析组别及各个水平的差异,而忽略了组别各个水平的交互作用;所以,此设计应是 2 因素的析因设计,因素之一是 2 个组别,因素之二是 4 个时间点,共 8 组,应为

析因设计的方差分析。原作者所用方法难以说明是时间作用还是处理组作用,也无法分析时间和处理组是否存在交互作用。

正确的表述举例如表 2 所示。

表2 A和B因素各水平下的WBC ( $\bar{x} \pm s, n=5$ )

B	A			合计	F值	P值
	A1	A2	A3			
B1	2.65 ± 0.86	2.70 ± 0.78	2.95 ± 0.88	2.77 ± 0.80	0.220	0.805
B2	1.85 ± 0.84	2.00 ± 0.53	2.10 ± 0.78	1.98 ± 0.69	0.180	0.837
合计	2.25 ± 0.91	2.35 ± 0.73	2.53 ± 0.91	-	0.377*	0.689*
F值	1.636	1.820	1.770	8.943*	5.861#	0.037#
P值	0.133	0.099	0.107	0.006*		

\*:主效应;#:交互效应。

分析:经析因方差分析,结果显示B因素各组间差异有统计学意义( $P=0.006$ ),B1比B2组WBC高。不同A水平间差异无统计学意义( $P=0.689$ ),A与B之间有交互作用( $P=0.037$ )。进一步分析单独效应,在固定A因素各水平条件下,B1和B2间比较均无统计学意义( $P$ 均 $>0.099$ ),在固定B因素各水平条件下,A1、A2、A3间差异也均无统计学意义( $P$ 均 $>0.805$ )。

综上,医学编辑应为作者提出以下指导性意见:

1)根据提供的设计方案,因素之间的交互作用被忽略,此设计应为析因设计方差分析。

2)若无交互效应,则可不必要对单独效应进行讨论,对于有统计学意义的主效应,可进行多重两两比较。若存在交互效应,则应分别讨论单独效应,即固定一个因素的某水平,另一个因素各个水平进行两两比较,不宜对主效应进行多重比较。

## 2 生存分析

生存分析是把结局变量和生存时间相结合起来的统计分析方法。常用到生存资料的分析,例如一些慢性疾病(高血压)、恶性肿瘤的预后和远期疗效观察以及术后患者生存质量状况及其影响因素等<sup>[5]</sup>。

**2.1 常见错误** 生存分析资料中,作者常忽略绘制生存曲线图、中位生存时间的95%可信区间等。

生存分析中2组之间生存率的比较,常被误用为2独立样本 $t$ 检验,此时应采用log-rank检验;但要注意其中一种特殊情形,即若2条生存曲线出现明显的交叉现象,运用log-rank检验可能会得出错误的结论,而此时Two-stage方法具有较好的检验效能,结论与实际更加符合<sup>[6]</sup>。所以,编辑人员应让作者及时提供生存曲线图,以避免作者错误运用。出现上述情形应建议作者根据生存曲线的实际情况采用正确的分析方法。

生存资料中常用的是Cox比例风险回归模型,它可分析多因素对生存期的影响及与生存时间有关的预后影响因素的分析,在随访研究资料分析中的应用比logistic回归模型具有明显的优势<sup>[7]</sup>。

**2.2 实例分析** 在《单侧肾上腺切除治疗肾上腺增生性高血压术后影响高血压复发因素的分析》临床研

究文章中,原作者回顾性分析2004年1月至2012年2月接受单侧肾上腺切除手术的84例肾上腺增生高血压患者的临床资料。对患者一般资料、术前肾上腺疾病相关生化检查结果、手术时间及病理类型等资料进行收集整理,采用多因素logistic回归分析肾上腺增生性高血压肾上腺切除术后高血压复发的影响因素。

分析:本研究长达8年之久,随访时间长短不一,高血压是否复发与随访时间必然有关;因此,此类资料应获得手术后至高血压复发的时间资料,采用生存分析方法来分析,例如log-rank检验、Cox回归模型分析等,而用logistic回归分析不是最合适的方法。

综上,医学编辑应为作者提出以下指导意见:

1)分析疾病预后的影响因素,结局变量和随访时间应被充分利用,采用COX比例风险模型,寻找影响生存结局和生存时间的危险因素和保护因素,而不是简单的logistic回归。

2)提供生存曲线图、中位生存时间及95%CI、生存率等;2组之间生存率的比较,应采用log-rank检验。生存曲线存在明显交叉时,Two-stage方法具有较好的检验效能。

## 3 Bland-altman法(BA法)

临床研究中的一致性检验,在医学期刊论文统计学方法运用错误分析中关于BA法的介绍很少见到。BA法是一种用2种方法结果的差值、均值及95%一致性(limits of agreement,LoA)绘制图形直观的方法,从而得出2种方法结果是否具有一致性<sup>[9]</sup>,对定量资料一致性的检验最好的选择采用BA法。

**3.1 常见错误** BA法在医学论文中经常只给出差值均数和95%LoA,而缺乏95%LoA可信区间的估计,以及缺少BA图形。LoA只是反映样本特征,而不能直接推断总体特征,所以应给出其置信区间<sup>[8]</sup>。

**3.2 实例分析** 《二维和三维超声测量方法对于单卵泡直径测量的一致性研究》临床研究中,方法是1位超声医生对438枚单卵泡发育的卵泡分别进行二维超声方法和三维超声方法的测量,通过BA法对二维直径和三维平均直径、二维直径和三维虚拟球径分别进行一致性评估。在结果中作者初投稿时只给出二维直径和三维平均直径对于单卵泡直径测量的一致性分析的差值、百分比、比值的BA图,分别有5.25%、6.16%、5.94%的点落在95%一致性区间以外,两者一致性不好;二维直径和三维虚拟球径对于单卵泡直径测量的一致性的分析,差值、百分比、比值的BA图,分别有5.25%、5.94%、10.96%的点落在95%一致性区间以外,两者一致性不好。结论:二维和三维超声测

量方法对于单卵泡直径测量并不一致。

分析:运用 BA 法进行一致性评价,一定要先给出临床可接受的允许误差限值,应该在结果中重点给出平均差值、95% LOA 以及 95% LoA 的 95% 可信区间结果。如果后一可信区间位于临床可接受的误差限值内,则可以认为 2 种测量方法具有一致性。如果测量数据行为良好(差值服从正态分布、无比例偏倚、无方差不齐),只需在二维直角坐标系中用横轴 X 表示同一个卵泡的二维直径(2D)和三维直径(3D-MDF 或 3D-DV)的平均值,纵轴 Y 表示两者的差值,得到差值 BA 图即可。

最终稿件结果正确表述为:在二维直径和三维平均直径 2 种测量方式对于单卵泡直径的测量的一致性分析中,差值的均数  $\bar{d} = -1.06 \text{ mm}$ , 95% LOA 为  $(-4.82, 2.70) \text{ mm}$ , 95% LOA 的 95% CI 为  $(-5.21, 3.08)$ , 临床可接受一致性区间为  $(-3.56, 1.44)$ 、 $(-5.21, 3.08)$ 、 $(-3.56, 1.44) \text{ mm}$ , 一致性不好;在二维直径和三维虚拟球径 2 种测量方式对于单卵泡直径的测量的一致性分析中,差值的均数为  $-0.07 \text{ mm}$ , 95% LOA 为  $(-3.11, 2.98) \text{ mm}$ , 95% LOA 的 95% CI 为  $(-3.43, 3.29) \text{ mm}$ , 临床可接受一致性区间为  $(-2.57, 2.43)$ 、 $(-3.43, 3.29)$ 、 $(-2.57, 2.43) \text{ mm}$ , 一致性不好。结论:二维和三维超声测量方法对于单卵泡直径测量并不一致<sup>[10]</sup>。

综上,医学编辑应为作者提出以下指导意见:

1) 请先给出临床可接受的允许误差限值。

2) 一定要给出平均差值、95% LOA 以及 95% LoA 的 95% 可信区间。如果后一可信区间位于临床可接受的误差限值内,则可认为 2 种测量方法具有一致性。如果测量数据行为良好(差值服从正态分布、无比例偏倚、无方差不齐),只需在二维直角坐标系中用横轴 X 表示 2 种方法的平均值,纵轴 Y 表示两者的差值,得到差值 BA 图即可。

#### 4 结束语

医学和医学期刊的发展,对医学编辑人员提出了

更高的要求,编辑人员不仅要做好审稿和编稿工作<sup>[11]</sup>,熟练掌握医学各个领域最新的发展动向和热点,还要熟练掌握医学论文常用的统计学方法,比如  $t$  检验、 $\chi^2$  检验、方差分析、秩和检验等简单的方法,还要熟悉高级生物统计方法,比如析因设计、重复测量、生存分析及 BA 法等,能及时发现稿件中统计学方法运用错误之处,最好能做到送审稿专家之前发挥自己的主观能动性,为作者提供有建设性的修改意见,以保证医学科研论文结果的准确性和科学性,并避免有价值的科研论文被退稿,从而推动医学科技期刊的发展。

#### 5 参考文献

- [1] 肖丽娟,孙茂民. 医学论文中统计学处理常见问题及应对措施[J]. 编辑学报,2010,22(6):500
- [2] 冉明会. 医学期刊编辑应重视摘要中统计学著录问题的审编[J]. 循证医学,2014,26(3):238
- [3] 方积乾. 卫生统计学[M]. 6 版. 北京:人民卫生出版社,2008:257
- [4] 李悦,朱凯,俞慧强. 析因设计资料单独效应分析及其 spss 程序实现[J]. 中国卫生统计,2009,26(6):643
- [5] 胡良平,郭辰仪. 用 SAS 软件实现析因设计定量资料的统计分析[J]. 药学服务与研究,2012,12(4):249
- [6] 李慧敏,韩栋,陈征,等. 生存曲线交叉时统计推断方法的比较和选择[J]. 中国卫生统计,2013,30(5):668
- [7] 李河,郭兰,孙家珍. cox 回归模型在临床医学科研中的价值[J]. 中华预防医学杂志,2011,11(1):51
- [8] 缪华章. 基于 Bland-Altman 法 LoA 可信区间的一致性评价[D]. 广州:南方医科大学,2014
- [9] BLAND J M, ALTMAN D G. Statistical methods for assessing agreement between two methods of clinical measurement [J]. Lancet,1986,1(8476):307
- [10] 朱颖春,贺方方,侯晓妮,等. 二维和三维超声测量方法对于单卵泡直径测量的一致性研究[J]. 医学研究生学报,2015,28(3):290
- [11] 贾泽军,尹茶,张建芬,等. 医学编辑成长经历分析及培养模式探讨[J]. 编辑学报,2013,25(1):79

(2015-10-22 收稿;2016-01-14 修回)

### 踏莎行·日暮闲思

赵大良/西安交通大学期刊中心

日暮西山,余晖渐褪,  
情牵两地心身累。  
清波碧水意悠然,  
孤枝摇曳如妃醉。

暂住滇西,扶贫不缀,  
虫鸣鸟啭声声碎。  
归期越近越思归,  
阑珊灯影催人寐。