

科技期刊编辑如何利用统计知识筛查学术不端行为

郑巧兰 林燕薇 洪悦民 杨江瑜

中山大学附属第三医院期刊中心《新医学》编辑部, 510630, 广州

摘要 论文的统计学质量是学术期刊质量控制的重中之重。本文介绍几种常用统计学检验方法的规则,结合案例分析,分享《新医学》在文章统计学质量把关方面的经验:1) t 检验和秩和检验中 $|t|$ 或者 $|Z|$ 若 < 1.96 , 对应的 P 值必然 > 0.05 , 且在相同样本量的情况下, $|t|$ 或者 $|Z|$ 越大, P 值越小。2) 四格表的卡方检验, χ^2 值若 < 3.84 , 则对应的 P 值 > 0.05 ; 在样本量相同时, χ^2 值越大, P 值越小。3) logistic 回归分析或 Cox 生存分析中, 自然对数的底(e)的回归系数(B)次方, 即 $\exp(B)$, 必须等于 R_0 或 R_H 值; R_0 或 R_H 值的 95% CI 的下限 = $\exp(B - 1.96E_s)$, 95% CI 的上限 = $\exp(B + 1.96E_s)$; 且 R_0 或 R_H 值的 95% CI 如果包含 1, 对应的 P 值必然 > 0.05 ; R_0 或 R_H 值的 95% CI 必然包含 R_0 或 R_H 值 $\exp(B)$ 。编辑人员应掌握基本统计知识, 并增强与作者的沟通, 刊发文章应有多次统计审稿, 并要求作者提供具体的统计值。通过上述规则和经验, 希望能够帮助科技期刊编辑初步筛查可能存在统计结果错误的文章, 以避免刊出因统计问题所致的结果造假类学术不端论文。

关键词 学术不端; 统计学; 科技期刊; 规则

How can editors of scientific journals use statistical method to screen academic misconduct//ZHENG Qiaolan, LIN Yanwei, HONG Yuemin, YANG Jiangyu

Abstract Statistical quality is the top priority of academic journal quality control. This article introduces the rules of several commonly used statistical tests combined with case analysis to share our experience in the statistical quality control of articles. 1) If $|t|$ or $|Z|$ in the t test and rank sum test was < 1.96 , the corresponding P value must be > 0.05 . And in the case of the same sample size, the larger $|t|$ or $|Z|$, the smaller the P value. 2) If the χ^2 value is < 3.84 in the Pearson chi-square test of the 2×2 table, the corresponding P value must be > 0.05 . When the sample size is the same, the larger the χ^2 value, the smaller the P value. 3) In logistic regression analysis or Cox survival analysis, the $\exp(B)$ (B is regression coefficient) must be equal to the R_0 (odds ratio) or R_H (hazard ratio) value. The Lower limit of 95% CI of the R_0 or R_H value = $\exp(B - 1.96E_s)$ (E_s is standard error), upper limit = $\exp(B + 1.96E_s)$. And if the 95% CI of R_0 or R_H value contains 1, then the corresponding P value must be > 0.05 . The 95% CI of the R_0 or R_H value must include the R_0 or R_H value ($\exp(B)$). Journal editors should master basic statistical knowledge, and enhance communication with authors. Published articles should have multiple statistical review, and the authors should be required to provide specific statistical values. We hope the above-mentioned rules and experience can help editors initially screen for articles that may have incorrect statistical results, and

avoid publishing academic misconduct papers with falsified data results.

Keywords academic misconduct; statistics; scientific journals; rules

Authors' address Journal Center of the Third Affiliated Hospital of Sun Yat-sen University, Editorial of Journal of New Medicine, 510630, Guangzhou, China

DOI:10.16811/j.cnki.1001-4314.2021.01.017

2019年,国家新闻出版署发布了《学术出版规范期刊学术不端行为界定》,指出论文作者学术不端行为包括但不限于剽窃、伪造、篡改、不当署名、一稿多投、重复发表等问题^[1]。学术不端问题不仅影响了科学的健康发展,也影响了社会对科学研究的信任。尽管《国家科技计划实施中科研不端行为处理办法(试行)》《学位论文作假行为处理办法》《发表学术论文“五不准”》《高等学校预防与处理学术不端行为办法》《关于进一步加强科研诚信建设的若干意见》等法律法规对学术不端行为的处罚有明确规定;但是由于利益驱使,学术不端行为仍屡见不鲜。对于学术不端行为,除了论文作者的自我约束外,也需要编辑工作者的“监督”^[2]。文献数据库的查重系统可以帮助编辑们识别剽窃、一稿多投、重复发表等问题,但仍较难识别伪造、篡改等行为。编辑若能掌握一定的统计知识,则可以从结果中筛查结果伪造、篡改等学术不端行为,减少“胡编乱造”的文章。

《新医学》从2014年开始,由流行病与卫生统计学专业毕业的编辑人员担任统计学编辑,对文章进行统计学把关,杜绝了学术造假的文章发表,掌握了一定经验,现分享如下。

1 统计规则帮助发现科学错误

在医学论文中, t 检验、秩和检验、卡方检验、logistic 回归分析、Cox 生存分析等都较为常见。这些统计分析的结果有一定的“规则”。

1.1 t 检验和秩和检验的常见规则和使用

$|t|$ 或者 $|Z|$ 若 < 1.96 , 对应的 P 值必然 > 0.05 , 且在相同样本量的情况下, $|t|$ 或者 $|Z|$ 越大, P 值越小^[3]。在《可切除原发性肝癌破裂出血急诊肝切除与 TACE 后二期肝切除的临床研究》^[4]一文表 2 中,术中

输血量2组间比较的 $t = 1.952, P = 0.017$,与“或者若小于1.96,对应的 P 值必然 >0.05 ”不符。上文同一表里术中出血量比较, $t = 2.547, P = 0.028$ 。《心房颤动患者射频消融术后运动耐力现状及影响因素分析》^[5]一文表1中 $VO_2\max$ 的吸烟与否2组间比较,作者给的 $t = 7.370, P = 0.012$;而BMI正常与否的2组间比较, $t = 2.671, P = 0.009$ 。在《影响聚乙二醇干扰素 $\alpha - 2b$ 治疗慢性乙型肝炎患者疗效的多因素分析》^[6]一文表1中,2组白蛋白比较, $t = 2.155, P = 0.134$;而丙氨酸氨基转移酶的 $t = 2.048, P$ 反而为0.043。上述结果均违反“在相同样本量的情况下, $|t|$ 或者 $|Z|$ 值越大, P 越小”。若要较准确地验证,可以采用 t 检验的计算公式(见式(1)和(2)),粗略估计 t 值和 P 值,可以发现上述 t 值和(或) P 值是完全不正确的。

方差齐性时,

$$t = \frac{X_1 - X_2}{\sqrt{S_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$\text{式中 } S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (1)$$

方差不齐时,

$$t' = \frac{X_1 - X_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (2)$$

式(1)和(2)可以录入到Excel表格中进行保存,每次只要输入2组的均数、标准差和样本量,即能得到估算的 t 值和 P 值,将其与作者的 t 值和 P 值进行比较。如果不计算,也可以在Excel中利用 t 分布的公式“T.DIST.2T”以及标准正态的公式“NORM.S.DIST”,根据作者给的 t 值和样本量(自由度)或者 Z 值,对应获得 P 值,核对该 P 值是否与作者给的 P 值相符。实际上不止 t 检验,方差分析等同样可以根据作者给的 F 值、样本量和组数(自由度),在Excel中利用公式“F.DIST.RT”核对 F 值与 P 值是否相符。

另外,作者若未提供均数和标准差等具体数据,文章只有条图、直方图,编辑同样可以根据图的高度以及误差线长短估计均数和标准差,再结合样本量,利用上述公式得到估算的统计值,用于结果的粗略判断。

1.2 四格表卡方检验的常见规则和使用

对于四格表的卡方(χ^2)检验,即2组率的比较。 χ^2 值若 <3.84 ,则对应的 P 值 >0.05 ;在样本量相同时, χ^2 值越大, P 值越小^[3]。《剖宫产术后肠梗阻的相关危险因素分析》^[7]一文表2中,2组产房中转剖宫产率比较 $\chi^2 = 11.520, P = 0.010$;然而2组手术时间 \geq

1h比例比较, $\chi^2 = 10.437, P = 0.001$ 。实际上,前者的 P 值是0.001,编辑若有基本的统计知识,则可避免这样的笔误。

《危重症早产儿胃肠外营养相关胆汁淤积的影响因素分析》^[8]一文表1中,同样可以发现这类问题。2组间性别比较的 $\chi^2 = 1.245, P = 0.141$;而抗生素使用比较 $\chi^2 = 1.352, P = 0.785$ 。经过计算,可以发现这篇文章表1里性别、抗生素使用、新生儿湿肺、新生儿肺透明膜病指标的 χ^2 值和 P 值均有误;表2的喂养困难组间比较 χ^2 值也有误;表3关于logistic回归分析部分更是错得离谱。编辑要是能掌握SPSS等软件的操作^[9],则根据作者给的样本量和率,可以很容易地得到真实的 χ^2 值和 P 值。

1.3 logistic回归分析和Cox生存分析的常见规则和使用

logistic回归分析或者Cox生存分析中,自然对数的底(e)的回归系数(B)次方,即 $\exp(B)$,必须等于 R_0 或 R_H 值(R_0 为优势比,即OR; R_H 为风险比,即HR)。 R_0 或 R_H 值的95%CI的下限 $= \exp(B - 1.96E_s)$,95%CI的上限 $= \exp(B + 1.96E_s)$,其中 E_s 为标准误。 R_0 或 R_H 值的95%CI如果包含1,那么对应的 P 值必然 >0.05 。 R_0 或 R_H 值的95%CI必然包含 R_0 或 R_H 值 $\exp(B)$ ^[10-11]。《乙型肝炎肝衰竭患者预后影响因素Logistic回归分析》^[12]一文表2里4个因素的 $\exp(B)$ 均不等于作者给的 R_0 值, R_0 值的95%CI也有误。《影响聚乙二醇干扰素 $\alpha - 2b$ 治疗慢性乙型肝炎患者疗效的多因素分析》^[6]一文表2中多因素logistic回归分析中, $\exp(B)$ 不等于 R_0 值,而且多个 R_0 值的95%CI不包含 R_0 值。《HBV相关慢加急性肝衰竭患者预后及其影响因素分析》^[13]表4中PT值的 $\exp(B)$ 亦不等于 R_0 值。《中国普外基础与临床杂志》27卷1期的《老年Ⅲ期右半结肠癌经腹腔镜全结肠系膜切除治疗的临床效果及预后分析》^[14]和《96例乳腺叶状肿瘤复发转移及预后相关因素分析》^[15]2篇文章均存在同样的问题。

以文献[15]的表2为例,至少有3项 R_0 和(或)其95%CI计算有误,见表1(原表不规范处未修改,表中最后一列为我们根据作者给的 B 和 E_s 计算的结果)。其中,统计对象是手术方式的扩大切除术者,根据作者给的 R_0 (0.435)和 P 值(0.007),可以认为扩大切除术局部复发的发生率较乳房全切除术低;根据作者给的 R_0 的95%CI(0.282,1.718),由于区间包含了1,则认为2种手术方式差异无统计学意义;但根据作者给的回归系数和标准误计算得到的 R_0 (95%CI)为[43.510(1.034,672.579)],认为扩大切除术的局

部复发率较乳房全切除术更高。那真实情况究竟是什么? 不得而知!

表1 影响PTB局部复发的logistic多因素分析结果

因素	β	标准误	Wald χ^2 值	P 值	OR(95.0% CI)	根据 β 和标准误计算的 OR(95% CI)
年龄 / 岁						
<40	1.857	0.897	4.287	0.038	6.404(1.104,31.104)	6.404(1.104,37.156)
40 ~ <50	0.128	0.776	0.027	0.869	1.136(0.248,5.195)	1.136(0.248,5.201) ¹⁾
≥ 50	—	—	—	—	1.000	
肿瘤类型						
良性	-4.706	1.297	13.175	<0.001	0.009(0.001,0.115)	0.009(0.001,0.115)
交界性	-3.064	1.191	6.615	0.010	0.047(0.005,0.482)	0.047(0.005,0.482)
恶性	—	—	—	—	1.000	
手术方式						
包块切除术	4.909	1.405	12.204	<0.001	1.355(0.863,2.127)	135.504(2.155,2127.707)
扩大切除术	3.773	1.397	7.299	0.007	0.435(0.282,1.718)	43.510(1.034,672.579)
乳房全切术	—	—	—	—	1.000	

注:¹⁾这个区间稍有差异,可能是由于数值修约的原因,因此暂不认为作者数据结果造假。

综上所述,已发表的文献中存在统计结果错误的文章并不少见。不管是有意学术不端行为,还是粗心大意导致的错误,最终得到的错误结果甚至是完全相反的结论。这类文章不但不能为临床医生提供有益的信息,甚至会给出错误的指导,造成临床事故!若编辑人员能掌握这些统计学基本知识,将大大减少这类严重错误的发生。

2 我们的经验

2.1 编辑人员应掌握基本统计知识,刊发文章应有多次统计审稿

编辑人员应有基本的统计知识,如掌握上述介绍的常用统计方法的几个小规则,则有助于初步筛查严重错误。如果不能把握文章统计方法或者结果的准确性,则应由具有统计专业知识的兼职人员等负责文章的统计把关工作。我们的统计审稿包含了统计初筛、统计初审、统计复审等环节,由卫生统计专业的全职编辑负责。从接收作者的投稿文章开始,统计审查就与科学性、创新性、不端检测等同时进行审核。统计初筛主要检查作者的研究设计是否合理,统计分析方法是否正确,初步核对统计数据有无问题。若存在较大问题,尤其是学术造假问题(如上述列举的多项错误同时存在),统计学编辑有权实行一步退稿。统计初审是在文章通过专家外审之后,进行文章统计方法和统计结果的详细审核,除了审查统计方法使用正确与否,还对部分数据计算核对,以核对作者方法与结果的统一性,以及结果的正确性。在统计初审时,为了方便作者修改,我们不仅提出问题所在,还给出建议的分析方法,告知作者如何修改等。这样可以避免作者的反复修改,节省时间,缩短了审稿周期。若文章的统计初审有问题,则待作者修改之后,会有统计二审这个环

节,主要是审核作者是否修改正确;若仍有错误,则会再次退给作者,返回的修改稿进行统计复审环节,直至没有统计问题。这些环节的存在,大大避免了投机取巧的作者编造结果来投稿,从而避免了学术不端行为。

2.2 要求提供具体的统计值

计算机技术的发展使得统计软件迅猛发展,各种统计学检验很容易通过统计软件实现,可以得到具体的统计值。但是,不少编辑部还允许作者只提供P值是否小于0.05。由于有统计学意义的结果更容易被接受,所以有一些作者可能不经过计算,便给出 $P < 0.05$ 。如《顺向、逆向开通技术在颈动脉串联闭塞致急性缺血性卒中治疗中的应用对比观察》^[16]一文中,顺向组一次取栓成功率30.0%(3/10),逆向组为52.9%(9/17),作者写 $P < 0.05$,实际上经过计算P值应该为0.424。若让作者给出具体统计值,可能可以提醒他勿造假。因此,为了减少此类情况,方便对比统计值,我们建议所有编辑部均要求作者提供具体的统计值,尤其是具体的P值。

2.3 增强与作者沟通

统计初筛时,如发现作者所使用的统计学方法是正确的,但个别结果与规则存在不合之处,则应提醒作者在修改时复核统计学结果。作者经过修改之后,若仍有统计结果错误,而作者坚称并非造假,可再次与作者沟通,详细了解其使用的统计学方法及处理步骤,提醒其注意是否抄写错误。如有的作者经常在2组定量资料的比较中,将方差齐性检验的结果当作组间差异比较的结果;定性资料的分析时,经常有作者写了用卡方检验,但是P值实际是从Fisher确切概率法处获得的。若能与作者多沟通其稿件细节,提醒作者核对注意事项,则可以帮助作者避免出现统计学错误。在沟通时,应注意尽量少使用术语,以简明易懂的语言说明

文章存在的明显统计学错误。如不加复核,将导致研究结果不可信,即使投寄他刊刊出也将严重影响个人的学术诚信。不少作者在与我们沟通后,重新审视研究的统计学处理步骤,仔细复核统计学数据,使文章结果更具科学性。如作者拒绝复核,也无法提供原始数据、统计学处理步骤以供复核,则对其文章数据的可信性存疑,应予退稿处理。

3 结束语

我国期刊数量近万种,且不断增加,其中医学期刊占了不小份额^[17]。数量不断增加,学术质量也要加以保证,而统计学审稿是期刊尤其是医学期刊质量控制的重中之重,是医学论文审稿中不可缺少的重要环节^[18]。但是近年来,科研造假事件频发^[19],统计学误用、乱用屡见不鲜^[20],文章质量堪忧。通过统计学审查,可以减少一部分存在“胡编乱造”学术不端的文章,提高刊发文章的质量。这就要求编辑必须提高自身的统计学素养^[21],在数据的把关中有所作为^[22]。希望以上规则以及经验可以帮助编辑同行们识别“造假”问题,提高对这类文章的警惕性,做好初步的质量把关。总之,编辑人员应该克服畏难情绪,努力学习统计学知识,掌握筛查统计结果造假等学术不端的能力。遇到困难时,可向统计学专家请教,同时从文章的研究设计、统计分析方法恰当与否、分析结果是否正确、结论是否合理等对文章的统计质量进行评价,使得文章具有较高的统计学质量。只有这样,才能避免发表存在数据造假问题的文章,提升文章的可信度和期刊的公信力,筑牢学术诚信和出版伦理底线。

4 参考文献

- [1] 全国新闻出版标准化技术委员会. 学术出版规范 期刊学术不端行为界定: CY/T 174—2019[S]. 北京: 国家新闻出版署, 2019
- [2] 印波. 科技期刊编辑部在科研失信行为查处中的主体责任[J]. 编辑学报, 2020, 32(4): 367
- [3] 方积乾. 卫生统计学[M]. 6版. 北京: 人民卫生出版社, 2008
- [4] 黎官印, 彭勇, 马海, 等. 可切除原发性肝癌破裂出血急诊肝切除与 TACE 后二期肝切除的临床研究[J]. 中国普外基础与临床杂志, 2020, 27(1): 50
- [5] 鲍志鹏, 孙国珍, 商淑华, 等. 心房颤动患者射频消融术后运动耐力现状及影响因素分析[J]. 中国全科医学, 2020, 23(32): 4083
- [6] 白一春, 邓霁红, 梅小平. 影响聚乙二醇干扰素 $\alpha-2b$ 治疗慢性乙型肝炎患者疗效的多因素分析[J]. 实用肝脏病杂志, 2019, 22(4): 492
- [7] 彭晶晶, 陶峰, 陈红波. 剖宫产术后肠梗阻的相关危险因素分析[J]. 安徽医学, 2020, 41(1): 18
- [8] 冯明华, 王霞, 赖春华. 危重症早产儿胃肠外营养相关胆汁淤积的影响因素分析[J]. 中国临床新医学, 2019, 12(11): 1226
- [9] 张文彤, 闫洁. SPSS 统计分析基础教程[M]. 北京: 高等教育出版社, 2004
- [10] 毛宗福, 余红梅. 18 Logistic 回归//方积乾. 生物医学研究的统计方法[M]. 北京: 高等教育出版社, 2007
- [11] 余红梅. 19 生存分析//方积乾. 生物医学研究的统计方法[M]. 北京: 高等教育出版社, 2007
- [12] 姜春华, 傅小凡. 乙型肝炎肝衰竭患者预后影响因素 Logistic 回归分析[J]. 国际医药卫生导报, 2017, 23(5): 655
- [13] 彭思璐, 刘冰, 孙宏, 等. HBV 相关慢加急性肝衰竭患者预后及其影响因素分析[J]. 实用肝脏病杂志, 2019, 22(4): 547
- [14] 王芳红, 张辉, 储怀祝, 等. 老年Ⅲ期右半结肠癌经腹腔镜全结肠系膜切除治疗的临床效果及预后分析[J]. 中国普外基础与临床杂志, 2020, 27(1): 72
- [15] 刘嫄灵, 冉冉, 涂刚. 96 例乳腺叶状肿瘤复发转移及预后相关因素分析[J]. 中国普外基础与临床杂志, 2020, 27(1): 60
- [16] 张斌升, 韩红星, 王浩, 等. 顺向、逆向开通技术在颈动脉串联闭塞致急性缺血性卒中治疗中的应用对比观察[J]. 山东医药, 2019, 59(30): 52
- [17] 沈敏, 侯鉴君. 我国医学期刊种类的现状与发展趋势[J]. 中国出版, 2008(9): 41
- [18] 彭敏宁, 昌兰. 统计学审稿: 医学期刊质量控制的重中之重[J]. 长沙铁道学院学报(社会科学版), 2007(1): 284
- [19] Scientific fraud; action needed in China[J]. Lancet, 2010, 375(9709): 94
- [20] 黄晨, 袁平戈, 张大志. 医学期刊来稿中有关统计学错误分析[J]. 现代医药卫生, 2013, 29(15): 2268
- [21] 邱芬, 曾令霞, 国荣. 统计学审查在医学论文审稿中的必要性[J]. 中国科技期刊研究, 2011, 22(4): 574
- [22] 吴红艳, 尹平. 医学期刊编辑应在数据的统计把关中有所作为[J]. 编辑学报, 2008, 20(5): 400

(2020-09-18收稿;2020-11-26修回)