

# 科技期刊编辑应培养对错误数据的敏感性

武建虎 李小萍 郭青

《武警医学》编辑部,100039,北京

**摘要** 结合实例分析,对科技期刊论文中错误数据的类型进行了归纳,并就编辑如何提高对数据的敏感性提出了建议。认为编辑不仅要有认真负责的态度、相关专业知识的学习和积累、掌握必要的统计学知识,还要对数据进行细致的复核,才能减少甚至消除科技论文中的数据错误。

**关键词** 科技期刊;编辑;错误数据;敏感性

**Sci-tech journal editors should develop sensitivity to wrong data//**  
WU Jianhu, LI Xiaoping, GUO Qing

**Abstract** By analyzing some actual cases, we summarize the types of wrong data in science papers, and propose some suggestions on developing sensitivity to wrong data for sci-tech journal editors. It is pointed out that editors should take serious responsibilities, learn and accumulate professional knowledge related to their fields, and master necessary knowledge of statistics. Besides, they should also take a careful review of data in the manuscripts in order to reduce such errors to the largest extent.

**Key words** sci-tech journal; editor; wrong data; sensitivity

**Authors' address** Editorial Board of Medical Journal of the Chinese People's Armed Police Forces, 100039, Beijing, China

科技期刊的文章中数据的重要性不言而喻,但由于种种原因文章中经常会出现一些错误的数字,直接影响了论文的准确性和科学性。作为一名科技期刊编辑,应该培养对错误数据高度的敏感性,才能避免或减少这类错误,把好论文发表的最后一关。笔者结合工作实践,对论文中出现的错误数据的类型作了归纳,并就编辑如何提高对错误数据的敏感性提出建议。

## 1 实例分析

**例 1<sup>[1]</sup>** 对象:选择择期行下肢骨科切开复位内固定手术患者 40 例,年龄 19~46 岁,体重 51~78 kg。

该例对性别构成没有介绍,对年龄的介绍也不完整,只给出了范围,没有给出均数和标准差。

一般来说,介绍临床研究对象时,必须对性别和年龄进行描述,对于计量资料还要求给出其集中趋势和离散趋势。反映集中趋势的指标主要有均数、中位数、众数等,反映离散程度的指标有极差、标准差、四分位数间距,不同的资料需要选用不同的指标。如资料属于正态分布,则选用均数和标准差,如是偏态,则需介

绍中位数和极差、四分位数间距。

**例 2<sup>[2]</sup>** Phadiatop Infant 和混合变应原 IgE 的判定标准相同,当荧光值达到 170 时(对应浓度为 0.35 kU/L)为阳性,进一步分级为:1 级  $\geq 0.35$  kU/L,2 级  $\geq 0.70$  kU/L,3 级  $\geq 3.5$  kU/L,4 级  $\geq 17.5$  kU/L,5 级  $\geq 50$  kU/L,6 级  $\geq 100$  kU/L。

该例分级标准含混不清,如  $\geq 0.35$  与  $\geq 0.7$  有重叠,前者包含了后者。这类错误还常见于对连续性数据如质量、长度、温度、时间、年龄等分组的情况<sup>[3]</sup>。例如年龄分组,“年龄/岁:0~10,10~20,⋯,50~60,>60”。这一表示不准确,也不科学。正确为:“年龄/岁:0~10,>10~20,⋯,>50~60,>60。”

**例 3** 作者在对某院抗高血压药物使用分析的原稿中说:“2008 年,钙拮抗药 10 410 万元、血管紧张素转换酶抑制药 1 770 万元、利尿药 575 万元、 $\beta$ -受体阻滞药 613 万元,血管紧张素 II 受体拮抗药 312 万元,合计约 1.4 亿元。”

三甲医院的总收入一般是 10 亿元左右,而该院仅高血压药物就达到 1.4 亿元,所以数据肯定有问题。经与作者沟通,原来数据统计方法有问题。实际数据为,钙拮抗药 297 万元、血管紧张素转换酶抑制药 50 万元、利尿药 2 万元、 $\beta$ -受体阻滞药 22 万元,血管紧张素 II 受体拮抗药 159 万元,合计 530 万元左右<sup>[4]</sup>。

**例 4** 一篇关于征兵制度改革后兵员质量分析与控制策略的原稿中,作者指出:“检出身体有边缘问题人员占新兵总数的 9.1%,其中,调换新兵占新兵总数的 0.4%,经全军终极鉴定医院明确诊断身体不合格并退兵占新兵总数的 18.05%。”

该例中,身体不合格退兵率 18.05%,此数字过大。经与作者沟通,原来是笔误。真实数字是 1.5%。

**例 5<sup>[5]</sup>** 结果:观察组气腹后 10、40、60 min 和放气后 10 min  $SO_2$ 、 $PeT_{CO_2}$  与对照组比较,差异显著 ( $P < 0.05$ ),见表 1。

该例表 1 的主要问题如下。

1)缺样本例数。对于需要作统计分析的数据,样本例数必不可少。本例中观察组 105 例和对照组 106 例没有在表中体现。

2)有笔误。麻醉前对照组  $PeT_{CO_2}$  的数值是  $6.83 \pm 3.23$ ,明显不同于观察组的  $36.38 \pm 3.90$ ,应该是笔误。

表1 两组不同时间节点 HR、MAP、SO<sub>2</sub> 和 PeTco<sub>2</sub> 变化比较(x±s)

组别		麻醉前	气腹后 10 min	气腹后 40 min	气腹后 60 min	放气后 10 min
观察组	HR(次/min)	80.21±11.64	78.51±12.08	79.62±10.61	80.40±9.66	95.65±8.64**
	MAP(mmHg)	96.0±10.50	83.4±7.35*	83.18±9.53*	90.15±7.95	89.55±7.65
	SO <sub>2</sub>	0.950±0.017	0.989±0.011*△	0.986±0.014*△	0.990±0.010*△	0.976±0.018*△
	PeTco <sub>2</sub> (mmHg)	36.38±3.90	41.85±7.73*△	39.00±1.65*△	41.25±2.40*△	46.50±4.28**△
对照组	HR(次/min)	81.09±12.24	112.27±10.47**	91.70±7.36**	89.65±9.45*	95.16±10.41*
	MAP(mmHg)	95.25±9.38	88.05±9.23*	88.28±8.63*	89.70±7.35*	89.33±10.95*
	SO <sub>2</sub>	0.957±0.017	0.948±0.015	0.951±0.011	0.936±0.009	0.960±0.013
	PeTco <sub>2</sub> (mmHg)	6.83±3.23	45.75±2.33**	46.50±3.98**	47.25±3.83**	43.58±3.15*

注:与本组麻醉前比较,\* $P<0.05$ ,\*\* $P<0.01$ ;与对照组比较,△ $P<0.05$ 。

3) 统计分析有误。根据2组样本的均数、标准差、例数可以推断2组总体均数之差的置信区间<sup>[6]</sup>。本例中,气腹后10、40、60 min,观察组HR与对照组的总体均数之差的95%的置信区间不包含0,说明2组之间有统计学差异。同样,气腹后10、40 min,观察组MAP均低于对照组,差异有统计学意义。

4) 同一指标均数和标准差的小数位数不一致。如观察组MAP麻醉前与气腹后10 min的数值。

此外,该表表题交代不清楚,栏目设计不合理,栏目也未采用“量/单位”这一标准化表示方式,量名称还有错。这些问题不在本文的研究范围,故不予讨论。

## 2 错误数据的类型

**2.1 与数据相关的表达有误** 本文例1、2,以及例5中1)、4)的错误都属于这种类型。这种错误与作者统计学基本知识缺乏有关。如果作者对计量资料的统计描述知识熟悉,那么例1的错误就可以避免。实践中,很多作者不管资料的分布类型,一味采用均数±标准差的形式,有时会出现标准差是均数几倍的现象。例2的错误说明作者对数学符号的使用不熟悉。

**2.2 数据本身的含义违背常理** 本文例3、4,以及例5中2)的错误属于这种类型。了解相关背景知识的人一眼就可看出数据有误。例3的作者收集数据时,不小心采用了错误的计算公式,导致数值变大,而事后也没有进行复核。例4以及例5中的某些错误可能是作者笔误造成的。

**2.3 数据之间的逻辑关系不符** 如果单看一个数据,不会发现有错误,但若与文章的其他部分综合考虑就可能发现错误。

1) 数据之间违反了一致性原则<sup>[7]</sup>。如方法中提到采用小鼠30只,但结果中却是25只,全文也没有交代怎么少了5只。理论上讲,这2个数字必须一致。此外,还要考虑文章中各部分之间、文字与图表之间,相应的数据是否一致。

2) 简单的逻辑关系不符。如例2中,分组之间出

现了数据重叠。此外,常见的错误还有构成比之和不等、率的计算有误等。

3) 统计推断有问题。如例5中3)的错误。此类错误不易发现,一般需借助计算器或统计软件才能发现,必要的话还需作者提供原始数据。

## 3 编辑如何提高对错误数据的敏感性

**3.1 要有认真负责的态度** 审稿时,固然要重视论文的创新性,但也不能忽视对数据的审核。如果数据出错,则论文表达的准确性就无从谈起,更不用说科学性和创新性了。实践中,如果一篇文章出现多起数据错误,则有理由质疑作者的科研态度和科研能力,即便是基金支持的论文也应该坚决退稿。本文所举的一些数据错误为什么发表前没被发现,就是由于责任编辑不负责、不认真。编辑工作是整个出版工作的核心<sup>[8]</sup>,作为编辑,一定要有认真、负责的态度,以防止一些有错误的论文发表。

**3.2 要有相关专业知识的学习和积累** 尽管多数科技期刊的编辑都具有从事相关专业的经历<sup>[9]</sup>;但在这个知识更新越来越快、交叉学科越来越多的时代,如果不能及时学习和积累知识,那么在审稿和加工过程中,面对一些新名词、新概念就会束手无策,即使数据有错误也未必能发现。实践中,对于不熟悉的知识,一定要与作者沟通,必要时还要向专业人士请教或通过网络查询,尤其对于论文中的核心数据,更不能出一丝差错。

**3.3 要掌握必要的统计学知识** 很多科研论文离不开统计分析,而目前论文中统计学分析错误的现象非常普遍<sup>[10]</sup>,从简单的对数据的统计描述到复杂的统计推断,几乎每个环节、每个部分都可能发现错误。作为编辑,掌握必要的统计学知识十分重要。建议编辑同人多看看如胡良平等<sup>[11]</sup>关于统计学误用方面的书,只有见的错误多了,才能练就一眼就会看出错误的本领。

**3.4 要对数据进行仔细的复核** 相比较前面3条,这条最为重要。对于数据较多的文章,审稿时一定要耐心,尤其是出现数据的地方,一定要放慢速度,理清全