

医学论文中易出现的统计学问题解析

姜 鑫

(南京医科大学医学书刊出版部,江苏 南京 211166)

摘要:医学论文中,统计学分析必不可少,统计学涵盖了研究的多个方面。而相关研究表明,医学期刊中统计学处理完全正确的比例极低。文章归纳了医学期刊中出现比较集中的统计学方法选取和描述问题,主要为抽样方法描述和统计学分析方法描述不详细、不准确等;统计描述中数值准确性的问题,主要涉及统计图表问题和数据缺失情况的处理。以期和论文撰写者及编辑同行进行交流,提高医学期刊相关表述的准确性,从而提高论文及期刊的质量。

关键词:医学论文;统计学;编辑

中图分类号:G232.2

文献标志码:A

文章编号:1671-0479(2016)06-480-003

doi:10.7655/NYDXBSS20160616

医学论文中,统计学分析必不可少,一项关于医学期刊的调查表明,95%的稿件均进行了统计学分析,而数据统计学处理完全正确的不足30%^[1]。统计学涵盖了研究的多个方面,有学者研究发现^[2],在医学论文中,统计研究设计、数据的表达与描述、数据的统计分析、统计分析结果的解释等各个环节,都或多或少存在问题,其错误率平均约为80%。统计学分析的正确与否、数据表达方式的准确与否都直接关系着论文的质量和学术水平。每一位医学研究工作者都应掌握一定的统计学知识。同时,这也对期刊的编辑提出了更高的要求:具备相应的统计学基础知识,能对大多数文章的统计学问题进行审核。医学期刊编辑掌握并按照统计学原理对来稿进行学术质量把关,无疑将使医学期刊的学术质量得到保证和提高^[3]。笔者结合编辑工作中遇到的实际案例,归纳医学期刊中出现比较集中的统计学方法选取和描述问题、统计描述中数值准确性的问题,以期和论文撰写者及编辑同行进行交流,从而提高医学论文中统计学方面的编校质量。

一、统计学方法问题

医学统计学方法的准确选择与应用,直接关系到研究结果的正确性,并最终决定科学研究的质量。由于有些作者对医学科研的统计学理论和方法的应

用缺乏深刻了解,在医学论文中错误应用统计学方法的现象时有发生。

(一)抽样方法描述

论文中样本量的统计描述包括样本抽样方法、样本量的计算、入选标准、排除标准等方面的介绍。在统计学中,把研究对象的全体称为总体,要了解总体的分布规律,在统计分析工作中,通常是从总体中抽取一部分个体进行观测,这个过程称为抽样。抽样方法分为随机抽样和非随机抽样,在医学研究中,为保证样本的代表性,通常选取随机抽样方法。在样本抽取过程中,每抽取一个个体,就是对总体进行一次随机试验,每次抽取的 n 个个体称为总体的一个容量为 n 的样本。随机抽样方法的要求是具有代表性和独立性。随机抽样方法主要有单纯随机抽样、系统抽样、分层抽样、整群抽样、多阶段抽样等。

医学论文中,抽样方法的描述务必要详细和准确,一些论文中抽样方法只简单介绍“采用随机抽样的方式选取”某研究对象,而不交待随机的方法具体是什么或根本不介绍抽样方法,均会降低文章的严谨性。如一篇文章调查对象介绍为“以南京市高校在校大学生为调查对象,随机抽取5所高校,共发放问卷618份”,文章中未介绍抽样方法的名称,只简单罗列了抽样的操作方法,从描述上看,抽取5所高校,容易认为是整群抽样方法,而整群抽样方法应该

收稿日期:2016-10-08

作者简介:姜鑫(1986-),女,重庆北碚人,硕士,编辑。

是抽取5所高校的所有学生为调查对象,而从样本量仅为618份来看,显然不是,这必然造成读者的困惑。另一篇文章研究对象介绍为“对南京医科大学本科大一、大二年级学生按班级采用随机整群抽样的方法,每个年级抽取三个班……分发问卷200份”,该介绍包含了抽样方法(随机整群抽样),具体操作方式(大一、大二每个年级抽取三个班)、样本数量(200份),就比较准确全面,应参考此类书写方式。

(二)统计学分析方法描述

统计学分析方法通常会在全文第一部分“对象与方法”中作为一个比较独立的部分予以专门介绍。有学者^[3]研究认为,这一部分应包含统计分析软件的介绍、数据表达方式的介绍、统计学方法的交代和检验水准的设定等至少四个方面的内容,在实际工作中,发现前三个方面的内容容易出现较多问题。

1. 统计分析软件介绍

介绍统计学软件时无软件版本信息,这是论文中容易出现的问题,如“采用SPSS统计软件进行数据分析”,缺少软件版本,应为“采用SPSS17.0统计软件进行数据分析”此类描述。另有一些文章无统计学软件的介绍,应予以避免。

2. 根据资料类型选取不同表达方式

正确判别统计资料的性质是合理选择统计分析方法的重要前提。统计资料一般分为定量资料、定性资料和等级资料三大类。定量资料又称计量资料,指通过度量衡的方法从每个观察单位上测得的指标,特点是用具体的数值表示,一般带度量衡单位,多为连续性资料,如年龄、身高、体重等。定性资料又称计数资料,为将全体观测单位按某种性质或特征分组,然后分别清点各组观察单位的个数,其特点是没有度量衡单位,多为间断性资料,如血型、性别等。等级资料也叫有序资料,指将观察单位按某种属性的不同程度或次序分成等级后分组计数的观察结果,如根据某种治疗方式,将患者分为治愈、好转、无效^[4]。

论文中根据资料形式采取不同的表达方式,符合正态分布的定量资料采用均数 \pm 标准差($\bar{x} \pm s$)表达,偏态分布的定量资料通常采用中位数(M) \pm 四分位间距(Q)表达。定性资料常用构成比、率表达。这部分内容常见的错误有:①数据表达方式未提及;②偏态计量资料误用均数 \pm 标准差表示;③采用全而杂的描述方式,如“数据以均数+标准差、率表示”,应分不同类型的数据进行分别描述。

3. 统计学分析方法选取的问题

论文中对所用统计学方法的交待,应准确、清楚,哪些数据采用何种统计学方法应该分别指出,这

样才会让读者一目了然,同时能够对作者采用的方法是否正确作出准确判断^[5]。例如一篇文章中“计量资料用均数 \pm 标准差表示。采用方差分析比较组间均数差异。两两比较采用 $P \leq 0.05$ 为差异有统计学意义”,未说明方差分析是针对定量资料多组间比较,且也未指出两两比较所采用的方法。相较之下另一篇文章“计量资料两组间比较采用 t 检验,多组间比较采用单因素方差分析,采用 q 检验进行两两比较”,则介绍得详细、准确。

另外较常出现的问题是统计学方法选取本身的错误,如等级资料的比较选取了卡方检验(通常为秩和检验),定量资料多组间的比较选取了 t 检验(应为方差分析)等。还有一类较容易出现的问题是涉及多组间的比较。多组间比较需要建立的一个思想是,首先需要进行多组间数据不全相同的一个统计学检验,这一步骤在定量资料中通常采用方差分析,在定性资料中通常采用卡方检验,当得到结果是 $P < 0.05$ 时,能得到的结论是多组间数据不全相同,而要进一步说明组间数据的差异,则需要再进行两两比较,数据的两两比较方法较多,可参考相关统计学书籍。论文中容易出现的问题是研究结果只计算了一个多组间比较的统计量,而在数据解释中却阐述其中的一组数据比另一组高,差异有统计学意义,这是不正确的,必须进行两两比较结果补充。

二、统计描述中数值的准确性

统计描述是对数据的直接处理和分析,目的在于通过一定方式的描述与整理,计算统计数据的特征值,进而发现其数量的规律性,为用样本统计量推断未知总体的参数提供充分的依据^[6]。论文的结果部分一般由大量数据构成,数据的准确、可靠是学术论文的核心^[7]。结果部分的展现形式通常是文字描述、表格和图相结合,其内容包罗万象,其中的问题也各有不同,以下将主要总结在医学期刊出现较多的统计图表和数据缺失问题。

(一)统计图表问题

统计图、表格是统计描述中常用的表达形式,其以形象直观、简单明了、清晰易懂的方式对数据进行描述,能将重要结果简洁清晰地表示出来^[8]。图、表在文中可独立支持论文结果,不再重复文字描述,否则保留一种即可。

1. 统计图的问题

文章中统计图的问题首先是统计图的滥用,在文章编辑加工过程中,很多作者的图会被删除,此类图的一个共同点是存在意义较低,图的内容通过文

字描述已能很清楚表达,再增加图只是耗费版面和重复,如描述男、女例数表达和比例,文字描述很简单就能完成,一些作者却要加上饼图再说明,就没有必要了。另一类较多的问题是图中项目的缺失,如经常使用的坐标图缺失横坐标或纵坐标的名称。

2. 统计表格的问题

表格的运用较多,统计学表格均应为三线表格,仅使用横线。表格中,应条目清楚、完整,论文中常出现的问题是缺少条目以及条目介绍不清楚;对于表格中每一项数据,应在表格栏目中标明度量衡单位;小数位数需要统一,此类问题较常出现,特别是当小数点后数字为0时容易被忽略,如当表格中率的小数位数均保留1位,一个数据为75%,此处就应修改为75.0%。

在表格中,要注意“0”、空白和“-”的运用,当数据为0时必须标0,而不能空白;当数据为不能测量时标“-”,而不能是0或空白;当数据没有测量或缺失时为空白,而不能标0。此外,描述相同内容的表格不要拆分,许多文章里面本应是描述一个内容的大表格拆分成了多个小表格而分别描述,这样会使文章内容显得复杂,读者在阅读时比较吃力,没能体现表格使数据一目了然的优势。

(二)数据缺失问题

论文中出现结果部分几个表格中的总调查样本数都不一样的情况,而文章中未对该情况作出合理解释。造成这一现象的原因是数据的缺失,如调查对象是100,每个对象有5个观察指标,而在数据分析时,发现某几个对象的调查结果不完整,只有3个或4个观察指标结果,在分析一个指标时总样本量为100,而分析另一个指标时总样本量不是100。对待这种有缺失值的情况,最好的办法是补齐资料,即再找到调查对象进行该指标的调查。但在论文撰写阶段再去寻找该名调查对象补齐数据可能性较小,如果缺失的对象数目较少,通常的处理方式为剔除该对象数据,即剔除那些数据不完整的对象,从而维护论文数据的完整性和一致性。这里就涉及到一个理想样本量和实际样本量的问题,如一篇文章中对对象介绍为“分发问卷200份,回收有效问卷197份,有效问卷回收率为98.5%”,可见该研究理想的样本量是200,而实际样本量为197,在结果分析中,均为这197例调查对象的数据。这篇文章中实际剔除了信息不全的对象,且也在文章中作出了描述。

而有一些文章中,却并未对缺失情况作处理或说明,从而造成计算构成比或率时分母的不一致而

影响研究结果统计描述的准确性。如一篇文章对象介绍为“对全体在校本科生519人进行调查”,而其中一项结果为“该学院现就读专业是高考第一志愿的人数仅47人,占调查人数的9.2%”,根据提供的数据计算百分比应为 $47/519 \times 100\% = 9.1\%$,和文中的结果不一致,如果文章直接刊出读者就会对文章结果的可信度产生质疑,经过编辑进一步追问作者,发现此指标的应答人数只有512人,该百分比是以512为分母计算得出,针对此种情况作者就应在对象中对数据有缺失的情况进行说明,并在计算百分比时提供分母的数值。

数据缺失的处理办法较多,可以参照统计学教材中相关章节内容,但应在论文中说明缺失值的具体处理方法,而不能出现结果中各个调查总样本数不一致而毫无说明的情况。

综上,医学期刊中,统计学方法的正确应用和准确表达直接关系到文章的学术水平。对于作者,统计学分析是数据处理的必备环节,也是文章能否被录用、文章学术价值高低的最主要考察部分;对于编辑,论文中统计学方法选用的正确性、统计学相关内容描述的准确性也是编辑加工环节的一大重点。这都需要作者和编辑不断进行统计学相关知识的学习和总结,从而减少论文中的统计学错误,确保论文的严谨性和科学性。

参考文献

- [1] 李霞,张印朋,闫苏平. 医学期刊作者来稿统计学应用与表述常见问题分析[J]. 中国科技期刊研究,2011,22(1):88-89
- [2] 胡良平. 医学统计实用手册[M]. 北京:人民卫生出版社,2004:1
- [3] 李旭清. 医学期刊编辑在解决论文统计学问题中的作用[J]. 科技情报开发与经济,2008,18(20):96-97
- [4] 陆守曾. 医学统计学[M]. 北京:中国统计出版社,2002:3
- [5] 冉明会,罗萍,邓丹. 医学期刊编辑应注意的几个统计学处理问题[J]. 编辑学报,2009,21(6):503-505
- [6] 张梅琳. 新编统计学[M]. 北京:立信会计出版社,2001:10
- [7] 朱久法,张彩虹. 学术论文中常见的一些数据问题及对编辑工作的要求[J]. 中国科技期刊研究,2010,21(4):546-548
- [8] 李少闻,雷晓梅,杨媛媛,等. 医学科技论文中统计图表的编辑加工技巧[J]. 技术与创新管理,2015(6):652-653,672