

## (十一) 医学论文中的数据表达与统计学处理

卢庆霞

(《中国脊柱脊髓杂志》编辑部 100029 北京市)

科技论文常以数据说话,因而不不管是临床研究、基础研究,还是流行病学研究,都离不开对数据的收集、整理和分析。一篇合格论文,要求数据准确无误,并对数据进行整理和分析,即选择正确的统计学方法进行分析并对统计结果进行合理的解释。

(1)数据表达。数据的表达包括数字和量的单位。在科技论文中应使用阿拉伯数字,采用国际单位制。但在论文中常遇到对数据的表达不当。(1)约数表达不当。在一些论文中,常可看到“大小约为 3mm 左右”、“增加了近 2 倍多”、“约 80% 以上的患者”等。显然这些描述是不够准确的,一是相同意义重复,如“大小约为 3mm”与“大小为 3mm 左右”意思相近,用任一种表达均可;另一是会引起误解或费解两种含义的表达,如“增加了近 2 倍多”是说明“增加了近 2 倍”还是“增加了 2 倍多”?应用“大多数、极个别、少部分、基本上”等不确定词,在科技论文中也是不科学的。(2)平均数应用不当。在对数据进行处理时,只有在合理分组的基础上对同质事物求均数才有意义,才能反映事物的特征;它适用于对称分布特别是正态分布的资料,能反映观察值的集中趋势。离开这两个条件,均数的意义不大,甚至会提供错误的信息。如有的论文中将不同年龄、不同身高、不同性别的某个测量数据放在一起计算平均值。这显然是不合理的。(3)有效数字表达或修约错误。实验观测或收集数据时均涉及到有效数字问题,即该量的精确度。在研究设计时使用什么量和单位一经确定,在整理数据过程中不能随意自定有效数字。如在测量径线的长度时,用精度为 0.1cm 的刻度尺,其测量结果的有效数字只能是小数点后两位数,如“3.10cm”或“31.0mm”,不能为了表示自己测量的数字精确,写成“3.100cm”或“31.00mm”;也不能省去后面的“0”,写成“3.1cm”或“31mm”。在进行统计学处理时,往往计算平均值,在计算的平均数后对其有效数字应进行修约,修约后的有效数字应与样本数字的位数相同,修约的原则是“四舍六入五看齐,奇进偶不进”,不能随意增加有效数字的位数以表示数据精确。(4)数值范围表示不规范。在书写百分数范围时省略前一数字的%,写成 50~60%;书写相同幂次的数字范围时只写后一个数值的幂次,写成  $1.5\sim 2.0\times 10^5$ 。这些都是不正确的。规范的写法应是 50%~60%、 $1.5\times 10^5\sim 2.0\times 10^5$ 。

(2)统计学处理。随着计算机技术的发展,医学统计软件包大量涌现,进行统计计算变得非常方便、快捷。但是,如何选适当的统计方法及对统计结果进行合理的解释却是每一位研究者都应该掌握的。因此,在整理数据前必须弄清楚所收集的资料是计数资料、计量资料,还是等级资料;是否为正态分析;样本数是否足够;有无为零的项等。进行统计学比较时是多组间比较还是两组间比较,是要作相关分析还是其它。显著性检验是双侧还单侧?这些都关系到统计方法的选择和统计学意义的判断,不能一律都用均数 $\pm$ 标准差( $\bar{x}\pm s$ )表示数据、一概用  $t$  检验,  $P<0.05$  就有显著性差异。其实在实际工作中符合  $t$  检验条件的资料并不多,特别是临床资料中,更多的应采用  $\chi^2$  检验或确切概率法。

医学论文中数据及其统计学处理方面存在的问题并不少见,一是不少作者只注重文章的专业内容,另是有些作者不了解数字用法的国家标准及统计学方法。因此,在今后的研究工作中,除重视研究内容论述外,对研究数据的收集、整理、分析、统计处理也不可忽视。这样才能提高论文的整体质量。