

# 综合类医学期刊已发表文献统计学报告质量及其影响因素分析

朱珈仪<sup>1</sup>, 刘元元<sup>2</sup>, 姜恬<sup>3,4</sup>

1. 四川大学华西第二医院编辑部、出生缺陷与相关妇女儿童疾病教育部重点实验室, 四川 成都 610041;
2. 四川大学华西公共卫生学院/四川大学华西第四医院流行病与卫生统计学系;
3. 四川大学期刊社; 4. 四川大学学报(医学版)编辑部

**摘要:目的** 对综合类医学期刊已发表文献中常见统计学报告质量进行调查,并分析其影响因素。**方法** 随机抽取综合医药卫生类 2 种 Medline 期刊和 6 种中文核心期刊,调查其 2010 年和 2020 年发表的所有原始研究文献共计 480 篇。基于 SAMPL 指南、ICMJE 建议、CONSORT、STROBE 及相关规范,采用自制调查表对文献的统计学报告质量进行评价,并分析其影响因素。**结果** 被调查的综合类医学期刊文献统计学报告质量的总分为  $40.0 \pm 8.6$ , 统计设计分值为  $43.2 \pm 6.8$ , 统计描述分值为  $67.0 \pm 24.3$ ,  $t$  检验分值为  $23.6 \pm 17.5$ , 方差分析分值为  $26.8 \pm 17.2$ ,  $\chi^2$  检验分值为  $24.2 \pm 13.7$ , 秩和检验分值为  $45.5 \pm 18.4$ , 相关分析分值为  $41.9 \pm 13.2$ , 回归分析分值为  $40.5 \pm 14.8$ , 生存分析分值为  $45.6 \pm 14.9$ 。单因素分析显示,2020 年发表的文献统计学报告质量总分值高于 2010 年 ( $35.8 \pm 7.4$  vs  $44.3 \pm 7.5$ ,  $t = -12.510$ ,  $P < 0.001$ ); 多水平模型结果显示,期刊收录体系、研究领域和出版年份是影响总分的主要因素,与 Medline 收录期刊相比,中文核心期刊总分值较低。与人为研究对象的文献相比,以其他为研究对象的总分值较低。与 2010 年文献相比,2020 年文献的总分值较高。**结论** 国内综合类医学期刊文献常见统计学方法报告质量在近十年有一定提高,但仍然有较大改进空间,建议加强相关报告规范推广力度。

**关键词:** 统计报告; 中文医学文献; SAMPL 指南

中图分类号: R195.1 文献标志码: A 文章编号: 1003-8507(2024)16-2918-08

DOI: 10.20043/j.cnki.MPM.202406181

## Analysis of the quality of statistical reporting and its influencing factors in publications from Chinese general medical journals

ZHU Jia-yi\*, LIU Yuan-yuan, JIANG Tian

\* Editorial Department, Key Laboratory of Birth Defects and Related Diseases of Women and Children,

Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, Sichuan 610041, China

**Abstract: Objective** To investigate the quality of statistical reporting in publications from general medical journals and analyze the influencing factors. **Methods** A random sample of 480 original research articles published in 2010 and 2020 was selected from two Medline journals and six Chinese core journals in the field of medicine and health. The quality of statistical reporting was evaluated using a self-developed checklist based on the SAMPL guidelines, ICMJE recommendations, CONSORT, STROBE, and other relevant standards, followed by an analysis of influencing factors. **Results** The overall quality score for statistical reporting in the surveyed general medical journals was  $40.0 \pm 8.6$ . The scores for different statistical aspects were as follows: statistical design ( $43.2 \pm 6.8$ ), statistical description ( $67.0 \pm 24.3$ ),  $t$ -test ( $23.6 \pm 17.5$ ), ANOVA ( $26.8 \pm 17.2$ ), chi-square test ( $24.2 \pm 13.7$ ), rank-sum test ( $45.5 \pm 18.4$ ), correlation analysis ( $41.9 \pm 13.2$ ), regression analysis ( $40.5 \pm 14.8$ ), and survival analysis ( $45.6 \pm 14.9$ ). Univariate analysis indicated that the overall quality score for statistical reporting in articles published in 2020 was higher than that in 2010 ( $35.8 \pm 7.4$  vs  $44.3 \pm 7.5$ ,  $t = -12.510$ ,  $P < 0.001$ ). Multilevel modeling showed that the indexing system, research field, and publication year were the main factors influencing the overall score. Compared to Medline-indexed journals, the overall scores were lower for Chinese core journals. Articles focusing on non-human subjects had lower scores compared to those involving human subjects. Additionally, the overall scores for articles published in 2020 were higher than those in 2010. **Conclusion** The quality of

基金项目: 中国高校科技期刊研究会 2021 年医学期刊专项基金(CUJS-YX-2021-2-6)

作者简介: 朱珈仪(1988—), 女, 硕士研究生, 研究方向: 流行病与卫生统计学

通信作者: 姜恬, E-mail: huanzijiang@hotmail.com

statistical reporting in general medical journals in China has improved over the past decade, but there remains substantial room for further improvement. It is recommended to enhance the promotion of relevant reporting standards.

**Keywords:** Statistical reporting; Chinese Medical literature; SAMPL guideline

统计学贯穿从研究设计到结果呈现的整个科学研究过程,统计学在科学数据的产生、分析及推断方面发挥关键作用,是获得客观结论的基础之一。统计学对于医学研究至关重要,恰当、准确地统计设计、分析和报告是保证医学研究质量,特别是研究结果重现性和可重复性的关键前提之一<sup>[1-3]</sup>。在中国<sup>[4-6]</sup>、印度<sup>[7-8]</sup>、韩国<sup>[9]</sup>等多个国家,包括基础医学<sup>[10]</sup>、临床医学<sup>[11-13]</sup>、口腔医学<sup>[14-15]</sup>、药学<sup>[16]</sup>等的医学多专业领域中,统计学设计、分析及报告不恰当的现象仍较为普遍,大部分研究并未严格遵守既有医学研究相关报告规范。这导致研究结果无法重复,甚至可能得出误导性结论,严重影响研究的可信度。

为了规范医学研究报告,提升不同类型研究报告的透明度、准确性和完整性,保证研究的可靠性,一系列医学研究报告指南被提出<sup>[17]</sup>。1979 年发布的国际医学期刊编辑委员会(International Committee of Medical Journal Editors, ICMJE)医学期刊学术著作实施、报告、编辑和发表建议对医学研究的实施与报告进行了规范,其中包括较为精简的统计内容相关要求<sup>[18]</sup>。1988 年,医学期刊论文的统计报告指南之详述与解读进一步提出 15 项统计要点<sup>[19]</sup>。自 1996 年以来, EQUATOR (Enhancing the Quality and Transparency Of health Research) network (www.equator-network.org)<sup>[17]</sup> 发布了一系列报告指南,包括临床试验报告统一标准(Consolidated Standards of Reporting Trials, 简称 CONSORT)<sup>[20]</sup>、增强流行病学中观察性研究的报告规范(Strengthening the Reporting of Observational Studies in Epidemiology, 简称 STROBE)<sup>[21]</sup>、诊断准确性研究报告规范(Standards for Reporting Diagnostic Accuracy Studies, 简称 STARD)<sup>[22]</sup>、动物研究之体内实验报告(Animal Research: Reporting in vivo Experiments, 简称 ARRIVE)<sup>[23]</sup>, 扩充了对不同类型研究的报告规范。随后,文献统计分析和方法学指南(Statistical Analyses and Methods in the Published Literature, 简称 SAMPL)进一步强调了如何更详细地报告基本统计方法和结果<sup>[24]</sup>。

在过去几十年里,中文医学研究数量大幅增加,但已发表文献的统计学报告质量尚未见报道。因此,本研究参考 SAMPL 和 ICMJE 建议、CONSORT、STROBE 等报告指南,采用自行设计的评分表和调查问卷,对 2010 年和 2020 年中国综合类医学期刊统计学报告质量及其影响因素进行了调查与分析,为探索

提高医学研究质量和透明度的可行途径提供实证参考。

## 1 资料与方法

**1.1 资料来源** 为了对中文医学期刊的统计报告质量进行全面分析,本研究纳入了综合医学与卫生领域的期刊。根据期刊索引、主办单位和文献作者地理分布,采用分层整群随机抽样的方法,从中文综合医学与卫生学科 8 种 Medline 期刊和 29 种核心期刊中抽取 2 种 Medline 期刊和 6 种核心期刊,纳入这 8 种期刊在 2010 年和 2020 年每期发表的原创研究。本研究样本量计算公式:

$$n = \left( \frac{Z_{\alpha/2}}{\delta} \right)^2 \times \pi(1 - \pi)$$

其中, $n$  为样本量, $Z_{\alpha/2}$  为标准正态分布  $\alpha$  下双尾  $Z$  值, $\delta$  为允许误差, $\pi$  为论文的统计学错误率。既往文献报道,医学论文中统计学相关错误率为 30% ~ 50%<sup>[15, 25]</sup>。取论文的统计学错误率  $\pi$  为 40%, $\alpha = 0.05$ , $\delta = 5\%$ ,代入公式得: $n = 369$ 。同时考虑可能出现的信息损失情况,拟调查不少于 480 篇文献。下载获取纳入本研究的全部文献全文。检索所得文献通过阅读题目、摘要,及浏览全文,按照纳入、排除标准进行筛选,筛选所得文献按照发表日期及页码排序,按照不同期刊进行编号。

**1.2 调查指标** 整合统计学报告相关规范,包括 SAMPL, ICMJE 医学期刊学术著作实施、报告、编辑和发表建议,医学期刊论文的统计报告指南:详述与解读, CONSORT, STROBE, NIHPG 及 CHAMP 等规范中有关统计学方法报告推荐条目作为调查内容,形成“医学研究常见统计学方法报告调查表”(简称“调查表”)。调查表包括统计设计、设计描述、 $t$  检验、方差分析、 $\chi^2$  检验、秩和检验、相关分析、回归分析、生存分析、其他共计 10 个部分。每题根据文献报告质量赋 0 ~ 2 分。对调查表总体,统计设计、统计描述、 $t$  检验、 $\chi^2$  检验、方差分析、秩和检验、相关分析、回归分析、生存分析各部分分别进行计分,最后均标准化加权为百分制得分,即分值范围为 0 ~ 100 分。

同时,收集文献基本信息。通过相关文献回顾和专家咨询,本研究纳入的可能影响文献统计学质量的指标包括:(1)研究领域,(2)发表年份,(3)期刊影响因子,(4)期刊收录体系,(5)期刊出版周期,(6)期刊主办单位,(7)期刊创刊年份,(8)有无审稿费,(9)有无版面费,(10)研究类型,(11)研究对象类型,(12)

主要研究结果为阳性,(13)第一作者单位类型、城市规模及地域,(14)通信作者单位类型、城市规模及地域,(15)是否为多单位合作,(16)是否有基金项目支持及级别,(17)是否报告利益冲突,(18)是否有医药类企业参与。

**1.3 质量控制** 对于医学文献统计学质量的调查,纳入的每篇文献按照调查表进行仔细审查。所有文献均由两位调查员分别独立审阅,以评估其统计学方法报告的情况。进行两轮审阅,每一轮由两位不同的调查员独立录入。如果有意见不一致的情况,采用讨论及第三方裁定的方式解决。采用 Epidata 3.1 进行双重数据录入和管理。

**1.4 统计分析** 本研究定量资料服从正态分布,故采用均数、标准差进行统计描述;对定性资料采用频数和率、构成比进行统计描述。文献统计学质量的单因素分析采用成组  $t$  检验或单因素方差分析。本研究调查了 8 种期刊的 480 篇文献,数据具有层次结构,表现在每种期刊的论文报告内容、格式存在一定程度的相似性,其统计学分析报告的状况也存在一定程度的相似性,即因变量调查表总分及各部分得分不独立,因此考虑采用多水平统计模型进行分析。水平 1 为文献,水平 2 为期刊。利用  $-2$  倍对数似然值 ( $-2 \log \text{likelihood}$ ,  $-2LL$ )、赤池信息准则 (Akaike information criterion, AIC) 和贝叶斯信息准则 (Bayesian information criterion, BIC) 来判断模型拟合优劣。

本研究使用 R 软件 (version 4.0.3) 及程序包

“MBESS” “corrplot” “dplyr” “car” “lme4” “Matrix” “MASS” “scatterPlotMatrix” 对数据进行分析。所有检验采用双侧检验,检验水准  $\alpha = 0.05$ 。

## 2 结果

**2.1 纳入文献基本情况** 所纳入杂志的主办单位分别为医院、学会、研究所、高校、高校和医院联合。2 本期刊不收取审稿费,6 本收取。不同专业(基础医学、临床医学、其他)的文献比例为 42.9% (206/480)、48.6% (233/480)、8.5% (41/480)。观察性研究比例为 48.8% (234/480),其中同一篇文献既包括观察性研究又包括实验研究的文献比较为 1.3% (6/480),试验性研究为 51.3% (246/480)。

51.9% (249/480) 文献以人为研究对象,26.0% (125/480) 为实验动物,14.2% (68/480) 为细胞株、菌株及病毒株,7.9% (38/480) 为其他,具体包括医用材料、药品、病理组织、器官、基因。在以人及人体组织、器官、细胞为研究对象的文献中,35.5% (102/287) 的研究报告了通过伦理学委员会审批,36.6% (105/287) 的研究报告了获得患者知情同意。在使用了实验动物的文献中,30.1% (43/143) 文献提及对于试验动物的处置符合动物伦理学或动物福利要求。1.5% (7/480) 文献结果为阴性,其余文献结果为阳性。42.1% (202/480) 文献为多单位合作。69.6% (334/480) 有基金项目支持。仅 6.0% (30/480) 文献报告了利益冲突。本研究纳入的文献基本情况见表 1。

表 1 本研究纳入的中文综合类医学期刊文献的基本情况 ( $n = 480$ )

Table 1 Basic information of Chinese general medical journals included in this study ( $n = 480$ )

项目	$n$ (%)	项目	$n$ (%)
研究领域		研究对象类型	
基础医学	206(42.9)	人	249(51.9)
临床医学	233(48.6)	实验动物	125(26.0)
口腔医学	3(0.6)	细胞/病毒/菌株	68(14.2)
公共卫生	9(1.9)	其他	38(7.9)
药学	2(0.4)	第一作者单位类型	
医学技术	27(5.6)	高校	98(20.4)
研究类型		研究所	19(4.0)
观察性研究	234(48.8)	医院	361(75.2)
实验研究	246(51.2)	其他	2(0.4)
发表年份		第一作者单位地域	
2010	240(50.0)	东部	295(61.5)
2020	240(50.0)	中部	120(25.0)
主要研究结果		西部	65(13.5)
阴性	7(1.5)	第一作者单位城市	
阳性	473(98.5)	一线	145(30.2)
基金项目级别		二线	223(46.5)
无	146(30.4)	三线及以下	112(23.3)
校级	11(2.3)	通信作者单位类型	
市级	18(3.7)	高校	90(18.7)

(续表)

项目	n(%)	项目	n(%)
省部级	162(33.8)	研究所	20(4.2)
国家级	140(29.2)	医院	368(76.7)
其他	3(0.6)	其他	2(0.4)
是否多单位合作		通信作者单位地域	
否	278(57.9)	东部	298(62.1)
是	202(42.1)	中部	122(25.4)
报告利益冲突		西部	60(12.5)
否	450(94.0)	通信作者单位城市	
是	30(6.0)	一线	150(31.2)
医药企业参与		二线	225(46.9)
否	477(99.4)	三线及以下	105(21.9)
是	3(0.6)		

**2.2 综合类医学期刊文献统计学报告质量** 全部 480 篇文献的总分为  $40.0 \pm 8.6$ 。统计设计部分分值为  $43.2 \pm 6.8$ , 统计描述部分分值为  $67.0 \pm 24.3$ ,  $t$  检验部分分值为  $23.6 \pm 17.5$ , 方差分析部分分值为  $26.8 \pm 17.2$ ,  $\chi^2$  检验部分分值为  $24.2 \pm 13.7$ , 秩和检验部分分值为  $45.5 \pm 18.4$ , 相关分析部分分值为  $41.9 \pm 13.2$ , 回归分析部分分值为  $40.5 \pm 14.8$ , 生存分析部分分值为  $45.6 \pm 14.9$ 。

**2.3 综合类医学期刊文献统计学报告质量影响因素**

**2.3.1 单因素分析结果** 对本研究所纳入的综合类医学期刊文献统计学质量总分进行单因素分析, 结果见表 2。不同期刊收录体系、出版周期、主办单位、是否收取审稿费的期刊文献, 以及不同研究领域、研究类型、研究对象类型、是否获得伦理委员会审批、是否获得知情同意、是否符合动物伦理要求、出版年份、第一作者单位类型、第一作者单位城市、通信作者单位类型、通信作者单位城市、是否报告利益冲突的文献统计学质量比较, 差异均有统计学意义。

表 2 综合类医学期刊文献统计学报告质量单因素分析结果( $n = 480$ )

Table 2 Univariate analysis results of statistical reporting quality in publications from general medical journals ( $n = 480$ )

相关因素	n	总分(分, $\bar{x} \pm s$ )	t/F	P
期刊收录体系			5.222	<0.001
Medline	120	$43.5 \pm 9.0$		
中文核心	360	$38.9 \pm 8.1$		
出版周期			6.979	<0.001
双月刊	120	$41.8 \pm 8.9$		
月刊	180	$38.1 \pm 8.1$		
半月刊	120	$39.9 \pm 8.1$		
周刊	60	$42.7 \pm 9.0$		
主办单位类型			3.923	0.004
高校	180	$41.6 \pm 8.8$		
医院	60	$40.0 \pm 5.8$		
研究所	60	$39.2 \pm 7.8$		
学会	120	$37.8 \pm 9.5$		
高校联合医院	60	$40.6 \pm 8.5$		
审稿费			-4.374	<0.001
无	120	$37.1 \pm 8.9$		
有	360	$41.0 \pm 8.3$		
版面费			1.248	0.213
无	60	$41.3 \pm 8.5$		
有	420	$39.8 \pm 8.6$		
研究类型			7.398	<0.001
观察性研究	234	$42.9 \pm 8.4$		
实验研究	246	$37.3 \pm 7.9$		
研究领域			20.308	<0.001
基础医学	206	$36.0 \pm 7.2$		
临床医学	233	$42.9 \pm 8.3$		
口腔医学	3	$35.2 \pm 8.9$		
公共卫生	9	$46.6 \pm 6.3$		

(续表)

相关因素	<i>n</i>	总分(分, $\bar{x} \pm s$ )	<i>t/F</i>	<i>P</i>
药学	2	36.3 ± 5.8		
医学技术	27	44.3 ± 9.0		
研究对象类型			32.412	<0.001
人	249	43.4 ± 8.3		
实验动物	125	36.0 ± 7.4		
细胞/病毒/菌株	68	36.2 ± 6.9		
其他	38	38.1 ± 7.9		
伦理审批 <sup>a</sup>			-4.999	<0.001
无	185	40.9 ± 8.1		
有	102	45.9 ± 8.2		
知情同意 <sup>b</sup>			-3.997	<0.001
无	182	41.2 ± 8.5		
有	105	45.2 ± 7.8		
动物伦理要求 <sup>c</sup>			-2.959	0.004
无	100	34.9 ± 7.0		
有	43	38.7 ± 7.3		
主要研究结果			-1.606	0.109
阴性	7	40.0 ± 8.6		
阳性	473	45.2 ± 5.7		
发表年份			-12.510	<0.001
2010	240	35.8 ± 7.4		
2020	240	44.3 ± 7.5		
第一作者单位类型			11.849	<0.001
高校	98	37.7 ± 8.5		
研究所	19	31.4 ± 8.3		
医院	361	41.1 ± 8.3		
其他	2	33.8 ± 13.3		
第一作者单位地域			1.372	0.255
东部	295	40.4 ± 9.0		
中部	120	38.9 ± 7.7		
西部	65	40.5 ± 8.3		
第一作者单位城市			6.449	0.002
一线	145	41.5 ± 8.8		
二线	223	38.5 ± 7.8		
三线及以下	112	41.1 ± 9.4		
通信作者单位类型			13.092	<0.001
高校	90	38.2 ± 8.7		
研究所	20	30.8 ± 6.6		
医院	368	41.1 ± 8.3		
其他	2	27.7 ± 4.6		
通信作者单位地域			1.662	0.191
东部	298	40.4 ± 9.1		
中部	122	38.8 ± 7.6		
西部	60	40.6 ± 8.1		
通信作者单位城市			4.610	0.010
一线	150	41.3 ± 9.0		
二线	225	38.8 ± 7.7		
三线及以下	105	41.0 ± 9.5		
是否多单位合作			0.077	0.939
否	278	40.1 ± 8.5		
是	202	40.0 ± 8.8		
基金项目级别			1.284	0.269
无	146	39.2 ± 8.6		
校级	11	44.8 ± 9.0		
市级	18	41.3 ± 7.6		
省部级	162	40.7 ± 8.5		
国家级	140	39.6 ± 8.7		
其他	3	39.0 ± 5.2		

(续表)

相关因素	<i>n</i>	总分(分, $\bar{x} \pm s$ )	<i>t/F</i>	<i>P</i>
报告利益冲突			-4.016	<0.001
否	450	39.6 ± 8.5		
是	30	46.0 ± 8.1		
医药企业参与			-1.282	0.201
否	477	40.0 ± 8.6		
是	3	46.4 ± 12.8		

注:a 涉及伦理审批的文献为 187 篇,b 涉及知情同意的文献为 187 篇,c 涉及动物伦理的文献为 143 篇。

**2.3.2 综合类医学期刊文献统计学质量总分多因素分析结果** 多水平模型结果(表 3)显示,与 Medline 期刊相比,中文核心期刊总分值较低。与人为研究对象的文献相比,以其他为研究对象的总分值较低。与 2010 年文献相比,2020 年文献的总分值较高。

**表 3** 综合类医学期刊文献统计学质量总分影响因素分析的多水平模型结果(*n* = 480)

**Table 3** Influencing factors analysis result of statistical quality score in general medical journals using multilevel model (*n* = 480)

变量	估计值(95% <i>CI</i> )	<i>t</i>	<i>P</i>
期刊收录体系			
Medline			
中文核心	-0.040 (-0.069 ~ 0.011)	-2.284	0.022
影响因子	-0.016 (-0.073 ~ 0.041)	-0.449	0.653
研究领域			
基础医学			
临床医学	0.019 (-0.017 ~ 0.056)	0.971	0.332
口腔医学	0.015 (-0.066 ~ 0.097)	0.347	0.728
公共卫生	0.053 (-0.003 ~ 0.111)	1.770	0.077
药学	0.002 (-0.088 ~ 0.094)	0.039	0.969
医学技术	0.019 (-0.022 ~ 0.063)	0.842	0.400
研究类型			
观察性研究			
实验研究	0.001 (-0.018 ~ 0.021)	0.143	0.886
研究对象类型			
人			
实验动物	-0.028 (-0.070 ~ 0.010)	-1.343	0.179
细胞/病毒/菌株	-0.039 (-0.080 ~ 0.004)	-1.757	0.079
其他	-0.034 (-0.063 ~ -0.003)	-2.166	0.030
主要研究结果			
阴性			
阳性	0.039 (-0.012 ~ 0.090)	1.472	0.141
发表年份			
2010			
2020	0.079 (0.066 ~ 0.092)	11.504	<0.001
第一作者单位类型			
高校			
研究所	-0.031 (-0.072 ~ 0.000)	-1.721	0.085
医院	0.002 (-0.014 ~ 0.018)	0.214	0.831
其他	-0.003 (-0.100 ~ 0.084)	-0.060	0.952
第一作者单位地域			
东部			
中部	-0.007 (-0.024 ~ 0.010)	-0.823	0.411
西部	-0.003 (-0.023 ~ 0.016)	-0.259	0.796
第一作者单位城市			
一线			
二线	-0.005 (-0.024 ~ 0.013)	-0.524	0.600
三线及以下	0.002 (-0.018 ~ 0.021)	0.227	0.820
是否多单位合作			
否			
是	-0.003 (-0.016 ~ 0.009)	-0.531	0.595

(续表)

变量	估计值(95% CI)	t	P
基金项目级别			
无			
校级	0.013 (-0.027 ~ 0.055)	0.585	0.559
市级	0.011 (-0.022 ~ 0.043)	0.639	0.523
省部级	0.011 (-0.004 ~ 0.028)	1.323	0.186
国家级	0.000 (-0.015 ~ 0.018)	0.039	0.969
其他	-0.044 (-0.121 ~ 0.031)	-1.109	0.267
报告利益冲突			
否			
是	-0.018 (-0.052 ~ 0.013)	-1.050	0.294
医药企业参与			
否			
是	0.023 (-0.052 ~ 0.099)	0.593	0.553

注: AIC = -1 213.70, BIC = -1 084.30, -2LL = 637.85。

### 3 讨论

统计学在科学数据的产生和分析以及推断方面发挥关键作用,是科学研究获得客观、可重复结果的基础。统计学方法的不恰当使用或不充分报告可能导致偏倚<sup>[3,25]</sup>。在过去几十年,中文医学研究数量大幅增加。本调查旨在调查 2010 年和 2020 年常见综合类医学期刊文献统计学报告质量,并衡量中文医学期刊遵守基本统计报告规范的情况,同时分析统计报告质量相关影响因素,为提高文献报告质量提供参考。

本研究结果表明,在一部分纳入的文献中,统计学报告的重要内容较模糊。Günel 等<sup>[13]</sup>选取 Thomson Reuters Clarivate Analytics 数据库中 20 种 SCI 或 SCI-E 索引的放射学期刊,随机抽取这 20 种期刊 2016 年至 2017 年发表的 157 篇文献进行审查发现,仅 10 篇文献没有统计错误。Amiri 等<sup>[26]</sup>纳入 99 篇 2019 年 Cochrane 收录的非药物康复随机对照试验文献进行评价发现,19.2% (19/99) 文献至少存在 1 种统计分析错误。Hassan 等<sup>[7]</sup>对 10 种印度高影响力医学期刊 2003 年( $n=588$ )和 2013 年( $n=774$ )发表的所有原始研究文献进行质量评价。结果显示,与 2003 年相比,2013 年存在统计错误文献比例并没有减少 [25% (80/320) vs. 22.6% (111/490),  $\chi^2 = 0.592, P = 0.4418$ ]。Jin 等<sup>[4]</sup>选择 10 种中国医学期刊,对 1998 年( $n=1335$ )和 2008 年( $n=1578$ )发表的所有原始研究文献进行全面分析。结果显示,统计分析中的误差/缺陷比例下降 [59.8% (545/1335) vs. 52.2% (664/1578),  $\chi^2 = 12.03, P < 0.001$ ],但一些问题仍然存在。上述研究结果显示医学论文统计质量普遍存在欠缺,有待提高。

根据本研究结果,对中文医学文献统计报告的完善提出如下建议。①作者在研究设计之初,咨询医学

统计学家并严格遵守 SAMPL 等指南和规范;②加强提升医学本科生和研究生的统计学实际运用能力以及论文撰写规范的教育培训,有条件的高校可邀请期刊编辑共同参与教学,更多采取案例教学法,以实际案例讲授如何更透明、更严谨的报告结果;③期刊应该加强对统计报告的严格审查;④鼓励医学期刊、生物统计学家等开展统计学科普,进一步提升作者、审稿人和编辑的数据素养。

本研究亦存在一些局限性。考虑到调查的可行性,从 37 种综合医学卫生类 Medline 或中文核心期刊中仅纳入 8 种期刊进行医学文献统计报告质量调查。相对较小的样本可能会导致偏倚,结果不应一概而论。由于调查的可行性和准确性,本研究缺乏对新的统计方法的调查和分析。此外,一些文献的一般信息不公开,如作者的教育背景,这使得本研究中纳入的相关因素不够全面。此外,需要注意的是,本研究评估的是文献的统计学报告质量而非文献的质量。尽管存在这些局限性,本研究将对提升中文医学文献报告的透明度、可信度提供一定参考,并提出了一些建设性和针对性较强的改进建议。

**利益冲突声明** 本研究不存在任何利益冲突

### 参考文献

- [1] Altman DG. Statistics in medical journals: some recent trends[J]. *Medicine Statistics*, 2000, 19(23): 3275-3289.
- [2] Horton NJ, Switzer SS. Statistical methods in the journal[J]. *New England Journal of Medicine*, 2005, 353(18): 1977-1979.
- [3] Altman DG. Poor-quality medical research: what can journals do? [J]. *JAMA: the Journal of the American Medical Association*, 2002, 287(21): 2765-2767.
- [4] Jin ZC, Yu DH, Zhang LM, et al. A retrospective survey of research design and statistical analyses in selected Chinese medical journals in 1998 and 2008[J]. *PLOS One*, 2010, 5(5): e10822.

(下转第 2942 页)

- start study[J]. *The Journal of Pediatrics*, 2020, 218: 28–34. e2.
- [22] Evlampidou I, Bagkeris M, Vardavas C, et al. Prenatal second-hand smoke exposure measured with urine cotinine may reduce gross motor development at 18 months of age [J]. *The Journal of Pediatrics*, 2015, 167(2): 246–252. e2.
- [23] Lee M, Ha MN, Hong YC, et al. Exposure to prenatal secondhand smoke and early neurodevelopment: Mothers and Children's Environmental Health (MOCEH) study[J]. *Environmental Health: a Global Access Science Source*, 2019, 18(1): 22.
- [24] Taylor AE, Carslake D, De mola CL, et al. Maternal smoking in pregnancy and offspring depression: a cross cohort and negative control study[J]. *Scientific Reports*, 2017, 7(1): 12579.
- [25] Roza SJ, Verhulst FC, Jaddoe VWV, et al. Maternal smoking during pregnancy and child behaviour problems: the Generation R Study[J]. *International Journal of Epidemiology*, 2009, 38(3): 680–689.
- [26] 刘贤, 郭程, 伊鹏, 等. 孕前及孕期母亲被动吸烟与儿童孤独症谱系障碍的关联分析[J]. *中国儿童保健杂志*, 2022, 30(4): 371–375.
- Liu X, Guo C, Yi P, et al. Association between maternal passive smoking exposure before/during pregnancy and the risk of autism spectrum disorders in offspring[J]. *Chinese Journal of Child Health Care*, 2022, 30(4): 371–375.
- [27] Margolis AE, Lee SH, Liu R, et al. Associations between prenatal exposure to second hand smoke and infant self-regulation in a New York city longitudinal prospective birth cohort[J]. *Environmental Research*, 2023, 227: 115652.
- [28] Walsh RA. Effects of maternal smoking on adverse pregnancy outcomes: examination of the criteria of causation[J]. *Human Biology*, 1994, 66(6): 1059–1092.
- [29] Yakel JL. Cholinergic receptors: functional role of nicotinic ACh receptors in brain circuits and disease [J]. *Pflugers Archiv: European Journal of Physiology*, 2013, 465(4): 441–450.

收稿日期: 2023-03-23

## (上接第 2924 页)

- [5] Wu SQ, Jin ZC, Wei X, et al. Misuse of statistical methods in 10 leading Chinese medical journals in 1998 and 2008 [J]. *The Scientific World Journal*, 2011, 11: 2106–2114.
- [6] Wang Q, Zhang B. Research design and statistical methods in Chinese medical journals[J]. *JAMA: the Journal of the American Medical Association*, 1998, 280(3): 283–285.
- [7] Hassan S, Yellur R, Subramani P, et al. Research design and statistical methods in Indian medical journals: a retrospective survey [J]. *PLOS One*, 2015, 10(4): e0121268.
- [8] Jaykaran G, Kantharia ND, Preeti Y, et al. Reporting statistics in clinical trials published in Indian journals: a survey [J]. *African Health Sciences*, 2010, 10(2): 204–207.
- [9] Park E, Cho M, Ki CS. Correct use of repeated measures analysis of variance[J]. *Korean Journal of Laboratory Medicine*, 2009, 29(1): 1–9.
- [10] Bahar B, Pambuccian SE, Barkan GA, et al. The use and misuse of statistical methods in cytopathology studies: review of 6 journals [J]. *Laboratory Medicine*, 2019, 50(1): 8–15.
- [11] Liang GP, Fu WL, Wang KF. Analysis of t-test misuses and SPSS operations in medical research papers[J]. *Burns Trauma*, 2019, 7: 31.
- [12] Blakley BW, Janzen B. Statistical considerations in otolaryngology journals[J]. *Otolaryngology – head and Neck Surgery: Official Journal of American Academy of Otolaryngology – Head and Neck Surgery*, 2017, 157(5): 745–747.
- [13] Günel karadeniz P, Uzabacı E, Atuş kuyuk S, et al. Statistical errors in articles published in radiology journals[J]. *Diagnostic and Interventional Radiology*, 2019, 25(2): 102–108.
- [14] Vähänikkilä H, Tjäderhane L, Nieminen P. The statistical reporting quality of articles published in 2010 in five dental journals [J]. *Acta Odontologica Scandinavica*, 2015, 73(1): 76–80.
- [15] Hannigan A, Lynch CD. Statistical methodology in oral and dental research: pitfalls and recommendations [J]. *Journal of Dentistry*, 2013, 41(5): 385–392.
- [16] Daniel-Corneliu L, Drugan T, Farcaş A, et al. Statistical reporting in pharmaceutical papers from Romanian journals [J]. *Farmacia*, 2015, 63(3): 394–401.
- [17] Enhancing the QUALity and Transparency Of health Research Network. EQUATOR Network: Reporting guidelines [EB/OL]. [2024-08-01]. <https://www.equator-network.org/reporting-guidelines/>.
- [18] International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals [EB/OL]. [2024-07-31]. <http://www.icmje.org/icmje-recommendations.pdf>.
- [19] Bailar JC3, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations [J]. *Annals of Internal Medicine*, 1988, 108(2): 266–273.
- [20] Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials [J]. *Annals of Internal Medicine*, 2001, 134(8): 657–662.
- [21] Von elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies [J]. *Annals of Internal Medicine*, 2007, 147(8): 573–577.
- [22] Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies [J]. *BMJ*, 2015, 351: h5527.
- [23] Kilkenny C, Browne WJ, Cuthill IC, et al. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research [J]. *PLoS Biology*, 2010, 8(6): e1000412.
- [24] Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines [J]. *International Journal of Nursing Studies*, 2015, 52(1): 5–9.
- [25] Scales CDJ, Norris RD, Preminger GM, et al. Evaluating the evidence: statistical methods in randomized controlled trials in the urological literature [J]. *The Journal of Urology*, 2008, 180(4): 1463–1467.
- [26] Amiri M, Kumbhare D. Randomized controlled trials in non-pharmacological rehabilitation research: a scoping review of the reporting of sample size calculation, randomization procedure, and statistical analyses [J]. *European Journal of Physical and Rehabilitation Medicine*, 2020, 56(6): 790–798.

收稿日期: 2024-06-13