

AMLC 与 PSDS 检测结果的比较研究^{*}■王福军¹⁾ 冷怀明^{2)**} 谭秀荣¹⁾

收稿日期: 2014-05-19

修回日期: 2014-08-07

¹⁾ 《遵义医学院学报》编辑部, 贵州省遵义市 563099²⁾ 《第三军医大学学报》编辑部, 重庆市沙坪坝区 400038 E-mail: lenghuaiming@gmail.com

摘 要 运用 CNKI 科技期刊学术不端文献检测系统和万方论文相似性检测系统对《第三军医大学学报》和《遵义医学院学报》605 篇来稿进行同步检测, 比较两种学术不端检测系统检测结果的差异性, 以及以文献量复制比 $\geq 10\%$ 、 $\geq 20\%$ 和 $\geq 30\%$ 三个等级比较两种检测系统检测结果的检出率及一致性。结果显示 AMLC 与 PSDS 的检测结果显示存在差异 ($Z = -8.15, P < 0.05$); 文献量复制比 $\geq 30\%$ AMLC 和 PSDS 检出率没有差别 ($P > 0.05$); 文献量复制比 $\geq 20\%$ 、 $\geq 10\%$ AMLC 和 PSDS 检出率有差别 (P 均 < 0.05); 三个等级的检测结果具有一致性 (P 均 < 0.001)。这一研究结果为编辑选择合适的检测系统提供了实证依据。

关键词 学术不端行为 AMLC PSDS 检出率

随着网络技术的不断发展, 各种学术不端行为冲击着科技期刊, 给科技期刊的发展带来了严峻挑战。各种学术不端检测系统也随之应运而生, 现已成为科技期刊编辑防范学术不端行为的必备武器。这些检测系统的研制成功为净化学术空气, 维护学术尊严, 打击和防范学术不端行为发挥了十分重要的作用。

目前, 国内运用最为广泛的学术不端检测系统主要有 CNKI 科技期刊学术不端文献检测系统 (AMLC)、万方论文相似性检测系统 (PSDS)、维普通达论文检测系统 (VTTMS)^[1]。AMLC 的国内用户已达 4600 余家, 其中期刊编辑部 4100 家, 研究生院 510 家; 而 PSDS 已经成为玛格泰克、勤云、三才等国内主流编辑平台的一个功能模块。AMLC 和 PSDS 现已成为科技期刊编辑甄别来稿是否存在学术不端行为的重要手段。对稿件文献量复制率等级与学术不端的关系^[2-3]及目前系统存在的问题^[4-6]等方面的研究广大科技期刊已进行了大量报道, 但对 AMLC 与 PSDS 检测结果是否存在差异及检出率等方面的研究鲜有报道。藉此, 本研究拟用 AMLC 与 PSDS 分别对《第三军医大学学报》和《遵义医学院学报》605 篇来稿进行同步检测, 比较两种检测系统检测结果是否存在差异以及以文献量复

制比 $\geq 10\%$ 、 $\geq 20\%$ 和 $\geq 30\%$ 三个等级检测结果的检出率及一致性, 探讨两种检测系统存在差异的文献量复制比等级及各等级检测结果的一致性, 以期科技期刊编辑选择合适的学术不端检测系统提供现实依据。

1 对象与方法

1.1 对象

以《第三军医大学学报》2014 年 4 月至 5 月和《遵义医学院学报》2013 年 9 月至 2014 年 5 月经编辑初审符合要求的 609 篇来稿为检测对象, 同步采用 AMLC 和 PSDS 对 609 篇来稿进行检测, 对两个检测系统的检测结果与原文进行逐一对比, 确认检测结果准确无误后剔除与个人学位论文重复超过 30% 的 4 篇稿件, 将余下的 605 篇稿件作为研究对象, 按文献量复制比区段对检测结果进行统计分析。

1.2 方法

按文献量复制比 $0\% - 10.0\%$ 、 $> 10.0\% - 20.0\%$ 、 $> 20.0\% - 30.0\%$ 、 $> 30.0\% - 40.0\%$ 、 $> 40.0\% - 50.0\%$ 、 $> 50.0\% - 60.0\%$ 、 $> 60.0\% - 70.0\%$ 、 $> 70.0\% - 80.0\%$ 、 $> 80.0\%$ 进行分段统计; 比较两种检测系统的总体检测结果是否存在差

* 基金项目: 遵义医学院硕士启动基金 (No. F-517)

** 通讯作者

异;以文献量复制比 $\geq 10\%$ 、 $\geq 20\%$ 和 $\geq 30\%$ 三个等级比较两种检测系统检测结果的检出率及一致性。

1.3 统计学处理

数据采用中位数与四分位间距(MP25, P75)表示,用SPSS13.0软件包对数据进行统计分析,两种检测系统的检测结果总体差异性比较采用Wilcoxon符号秩和检验,文献量复制比 $\geq 10\%$ 、 $\geq 20\%$ 和 $\geq 30\%$ 三个等级检测结果的检出率比较用McNemar检验、一致性比较用Kappa检验, $P < 0.05$, $P < 0.001$ 表示差异有统计学意义。

2 结果

605篇来稿均经两种检测系统进行同步检测,我们对两种检测系统的检测结果进行统计分析后得到以下结果。

2.1 文献量复制比各区段的论文篇数及构成比

对用两种检测系统检测的《第三军医大学学报》和《遵义医学院学报》605篇来稿按文献量复制比区段统计其论文篇数及构成比,结果见表1。

表1 两种学术不端检测系统检测来稿复制比各区段的论文篇数及构成比($n=605$)

复制比区段(%)	AMLC	构成比	PSDS	构成比
0~10.0	361	59.7	419	69.3
>10.0~20.0	121	20.0	91	15.0
>20.0~30.0	67	11.1	46	7.6
>30.0~40.0	26	4.3	23	3.8
>40.0~50.0	13	2.1	13	2.1
>50.0~60.0	9	1.5	4	0.7
>60.0~70.0	2	0.3	0	0.0
>70.0~80.0	3	0.5	4	0.7
>80.0	3	0.5	5	0.8

2.2 两种检测系统检测结果比较

将两种检测系统检测结果的总体差异性比较采用Wilcoxon符号秩和检验,结果显示两种检测系统的检测结果存在差异,检测结果AMLC高于PSDS, $Z = -8.15$, $P < 0.05$ 差异有统计学意义(结果见表2)。

表2 两种学术不端检测系统检测结果比较(MP25, P75 $n=605$)

分组	文献量复制比(%)
PSDS	4.76(0.00, 12.50)
AMLC	7.10(3.00, 17.35)
Z	-8.150
P	0.000

注:PSDS与AMLC比较, $P < 0.05$

2.3 以文献量复制比 $\geq 10\%$ 、 $\geq 20\%$ 和 $\geq 30\%$ 三个等级比较两种检测系统检测结果的检出率及一致性

作McNemar检验和Kappa检验,结果显示文献量复制比 $\geq 30\%$ AMLC和PSDS检出率没有差别, $P > 0.05$ 差异无统计学意义;文献量复制比 $\geq 20\%$ 、 $\geq 10\%$ AMLC和PSDS检出率有差别, P 均 < 0.05 差异有统计学意义;Kappa值分别为0.492,0.571,0.458, P 均 < 0.001 ,差异有统计学意义,三个等级的检测结果具有一致性(结果见表3)。

表3 文献量复制比 $\geq 30\%$ 、 $\geq 20\%$ 和 $\geq 10\%$ 三个等级检测结果的检出率及一致性比较($n=605$)

分组	$\geq 30\%$		$\geq 20\%$		$\geq 10\%$	
	稿件 (篇数)	百分比 (%)	稿件 (篇数)	百分比 (%)	稿件 (篇数)	百分比 (%)
PSDS	50	8.3	98	16.2	189	31.2
AMLC	58	9.6	124	20.5	248	41.0
P	0.322		0.004		<0.001	
Kappa	0.492		0.571		0.458	
P	<0.001		<0.001		<0.001	

注:PSDS与AMLC比较

3 讨论

学术不端检测系统现已广泛应用于编辑实践中,成为科技期刊编辑甄别来稿是否存在学术不端行为的必备武器,对科技期刊防范和打击各种学术不端行为具有十分重要的作用。现国内常用的两种检测系统主要是AMLC和PSDS,两种检测系统的优缺点广大科技期刊编辑在实践应用中已认真总结^[7-10]。这些实践经验对编辑选择适合自己期刊的检测系统具有十分重要的指导意义。

本研究对目前国内最常用的两种学术不端检测系统进行了比较研究。通过两种检测系统对《第三军医大学学报》和《遵义医学院学报》经编辑初审符合要求的605篇来稿进行了同步检测,AMLC与PSDS文献量复制比小于等于30%的分别为90.8%和91.9%;>30%~50%的分别为6.4%和5.9%;大于50%分别为2.8%和2.2%。从表1可看出对于文献量复制比较高的稿件,两种检测系统的差异不大,直观数据可初步判断AMLC的文献量检出率要高于PSDS。由于两种检测系统的检测结果为非正态分布资料,故对其文献量复制比进行了Wilcoxon符号秩和检验,两种检测系统的检测结果存在统计学差异($Z = -8.15$, $P < 0.05$)。从表2可看出

AMLC 的文献量检出率高于 PSDS, 似乎与初步判断相吻合。对文献量复制比 $\geq 30\%$ 、 $\geq 20\%$ 和 $\geq 10\%$ 三个等级比较两种检测系统检测结果的检出率及一致性进行统计分析, 从表 3 可看出, 两种检测系统的检测结果在文献量复制比 $\geq 30\%$ 时无统计学差异 ($P > 0.05$), 说明以 30% 为分界两种检测系统的检出率无差异; 文献量复制比 $\geq 20\%$ 和 $\geq 10\%$ 两个等级的检测结果则有统计学差异 (P 均 < 0.05), 说明以 20%、10% 为分界两种检测系统的检出率有差异, 两个等级均显示 AMLC 的文献量检出率要高于 PSDS; 通过 Kappa 检验对三个等级的一致性进行分析, 结果显示两种检测方法呈正相关 (P 均 < 0.001), 说明 AMLC 检测文献量复制比高的稿件 PSDS 同样也高。而导致 AMLC 文献量复制比检出率高于 PSDS 的原因可能在于两种检测系统的数据库比对范围不同, AMLC 除以公开发行的科技期刊外, 还以学位论文、内部刊物、专利数据库、重要报纸全文数据库、互联网资源等为比对资源, AMLC 还增加英文数据比对资源 (德国 Springer、英国 Taylor&Francis 期刊数据库等), 而万方仅有中国学术期刊数据库 (CSPD)、中国学位论文全文数据库 (CDDb)、中国学术会议论文数据库 (CCPD)、中国学术网页数据库 (CSWD) 4 种数据库比对资源, AMLC 的数据库比对源多于 PSDS。还有一个可能的原因是期刊独家授权的影响, 由于部分期刊独家授权给了某家数据库, 另一家的数据库则无法将其作为学术不端检测系统的比对资源, 如中华医学系列杂志独家授权万方, 对于拼凑的论文两种检测系统基本都能检出, 说明两者的检测结果具有一致性。对文献量复制比没有特定要求的期刊, 如《第三军医大学学报》就没有以特定的文献量复制比作为判定学术不端的依据^[11], 编辑可任选一种检测系统作为自己期刊的检测工具, 如果贵刊以 $< 20\%$ 或 $< 10\%$ 作为用稿的标准, 则建议选用 AMLC, 其余检测学术不端的方法及手段可作为补充。我们又对两种检测系统检测的 $\geq 30\%$ 稿件进行了分析, 部分稿件是与作者自己的学位论文重复, 其中 PSDS 的数据库未收录或收录不全该作者所毕业院校的学位

论文, 故不能检出该类稿件, 当然, AMLC 也有同样的情况, 因此, 学位论文的收录不全也是导致两种检测系统检测结果存在差异的原因之一。学位论文的再发表我们科技期刊编辑应尽量选择两种检测系统进行检测, 做到心中有数, 确保自己刊物所发稿件不存在学术不端行为, 对实践过程发现的它刊学术不端行为, 应主动函告相关刊物编辑部, 以共同防范和打击科技期刊中的各种学术不端行为。

综上所述, 本研究总体结果可信, 对于存在的不足及不同学科的论文进行检测是否会得到同样结果, 我们将在今后的研究中进一步探讨。检测系统所提供的结果只是为编辑提供参考, 其结果并不能判定该论文是否存在学术不端, 编辑在运用好各种学术不端检测系统的同时应充分发挥编辑的主观能动性。本研究为编辑正确选择适合自己刊物的学术不端检测系统提供了实证依据, 以更好地防范科技期刊来稿中的学术不端行为和提高稿件质量。

参考文献

- 1 张旻浩, 高国龙, 钱俊龙. 国内外学术不端文献检测系统平台的比较研究. 中国科技期刊研究, 2011, 22(4): 514-521
- 2 杨柱星, 林荣慧, 胡榜利等. 医学论文学术不端的特点及判定. 中国科技期刊研究, 2011, 22(5): 697-701
- 3 郑雨田, 齐烨, 李娜等. 医学文章文字重合度 20% - 30% 的内容对修改的提示. 中国科技期刊研究, 2014, 25(5): 645-647
- 4 谢文亮, 李俊吉, 张宜军. 期刊学术不端文献检测系统误检分析. 中国科技期刊研究, 2013, 24(6): 1126-1129
- 5 贺晋瑜, 杜学梅, 李敏生等. 科技期刊学术不端文献检测系统应用探讨. 中国科技期刊研究, 2013, 24(4): 777-779
- 6 杨晨晨. 运用学术不端文献检测系统检测医学论文存在的问题及对策. 编辑学报, 2014, 26(1): 42-44
- 7 秦小川. 《中医学报》应用 AMLC 检测结果分析和体会. 中国科技期刊研究, 2013, 24(6): 1122-1125
- 8 蔡丽枫, 刘冰, 游苏宁等. 中华医学会系列杂志论文相似性情况分析. 编辑学报, 2010, 22(3): 208-210
- 9 宋爽. 论文相似性检测服务在编辑工作中的使用与思考. 第十届 (2012) 全国核心期刊与期刊国际化、网络化研讨会论文集, 2012: 77-80
- 10 王胜海, 钟瑛. 利用技术手段遏制科研不端行为的实践与思考. 中国科技资源导刊, 2010, 42(5): 11-14, 48
- 11 吴培红, 冷怀明, 汪勤俭等. 防范科技论文一稿多投和重复发表. 编辑学报, 2009, 21(5): 434-436