

科技期刊总被引频次和影响因子构成中的自引率比较

——兼谈影响因子的人为操纵倾向*

刘雪立¹⁾ 魏雅慧^{1,2)} 盛丽娜¹⁾ 方红玲¹⁾ 王燕¹⁾ 付中静¹⁾ 郑成铭¹⁾ 董建军¹⁾

1)新乡医学院期刊社,河南省科技期刊研究中心;2)新乡医学院管理学院:453003,河南新乡

摘要 以2015年自然科学版JCR(SCI-JCR)收录的8778种期刊为研究对象,计算各期刊总被引频次构成中的自引率(self-citation rate for total citations, $R_{SC,TC}$)和影响因子构成中的自引率(self-citation rate for impact factor, $R_{SC,IF}$),全方位比较二者差异程度,揭示不同国家和地区人为操纵期刊影响因子的倾向。结果显示:8778种期刊集合 $R_{SC,TC}$ 为0.072,集合 $R_{SC,IF}$ 为0.112, $R_{SC,TC}$ 平均值为0.093, $R_{SC,IF}$ 平均值为0.132。排除17种无 $R_{SC,IF}$ 者,8761种期刊中, $R_{SC,IF} > R_{SC,TC}$ 者6954种(79.2%), $R_{SC,TC} > R_{SC,IF}$ 者1599种(18.2%), $R_{SC,IF} = R_{SC,TC}$ 者208种。Wilcoxon秩检验结果是, $Z = 59.836, P < 0.001$, 2组差异有统计学意义。SCI来源期刊中 $R_{SC,IF}$ 显著高于 $R_{SC,TC}$, 捷克、印度、日本、匈牙利和加拿大等国家期刊尤为突出。 $R_{SC,IF}$ 过度高于 $R_{SC,TC}$, 怀疑有人为操纵的可能。

关键词 科技期刊;总被引频次;影响因子;自引率;人为操纵
Comparisons of the self-citation rates for total citations and impact factors of journals: manipulation of journal impact factor//LIU Xueli, WEI Yahui, SHENG Lina, FANG Hongling, WANG Yan, FU Zhongjing, ZHENG Chengming, DONG Jianjun

Abstract This paper explore the differences between the self-citation rate for total citations ($R_{SC,TC}$) and self-citation rate for impact factors ($R_{SC,IF}$) in the SCI-sourced journals, and reveals the manipulation of journals impact factor in different countries and regions. The study objects include 8778 journals indexed in the JCR (SCI-JCR) in 2015, the $R_{SC,TC}$ and $R_{SC,IF}$ were calculated, and the differences between the $R_{SC,TC}$ and $R_{SC,IF}$ were compared. In the 8778 journals, the aggregative $R_{SC,TC}$ and aggregative $R_{SC,IF}$ are 0.072 and 0.112, respectively, the average $R_{SC,TC}$ and average $R_{SC,IF}$ were 0.093 0.132, respectively. In the 8761 journals (except 17 journals without $R_{SC,IF}$), 6954 (79.2%) journals have an $R_{SC,IF} > R_{SC,TC}$, and 1599 (18.2%) journals have an $R_{SC,TC} > R_{SC,IF}$, 208 journals have an $R_{SC,IF} = R_{SC,TC}$. Wilcoxon rank test indicated that $Z = 59.836, P < 0.001$, showing the difference between the two groups is statistically significant. $R_{SC,IF}$ in SCI-sourced journals is significantly higher than that of $R_{SC,TC}$, especially in the journals of Czech Republic, India, Japan, Hungary, and Canada. $R_{SC,IF}$ is obviously higher than $R_{SC,TC}$, which is suspected to be manipulated.

Keywords science journal; total citation; journal impact factor; self citation rate; manipulation

First-authors' address Periodicals Publishing House/ Henan

* 国家社会科学基金项目(15BTQ061)

Research Center for Science Journals, Xinxiang Medical University, 453003, Xinxiang, Henan, China

DOI:10.16811/j.cnki.1001-4314.2017.06.029

《期刊引证报告》(Journal Citation Reports, JCR)于1975年创立,影响因子被确定为JCR中评价期刊影响力的重要指标之一^[1-2]。近20年来,影响因子被广泛应用于学术期刊评价^[3]。澳大利亚Bradshaw等^[4]指出:无论你喜欢还是厌恶,用文献计量学指标评价学术期刊及相关研究绩效它就在那里;无论正确与否,学术组织评价申请人业绩、学者选择期刊投稿、学术期刊选择出版公司合作它都无所不在。

随着影响因子在期刊评价中地位的不断升高和应用的越来越普遍,人为操纵期刊影响因子^[5-7]的行为便自然而然产生了,而人为操纵影响因子最直接的手段就是提高期刊的自引。

希腊学者Falagas等^[8]总结了人为操纵影响因子的10大行为,第1和第3条都涉及期刊自引。国内学者Yu等^[9]分析了通过要求作者引用目标期刊操纵影响因子的现象,Bartneck等^[10]还研究了通过提高自引人为操纵h指数的现象。

国内最早研究期刊自引的是王崇德^[11]和罗式胜^[12],主要是介绍自引的概念、类型和计算方法。近年来,人们开始关注自引与影响因子的人为操纵^[13-16]。1999年版《中国科技期刊引证报告》把他引总引比(2005年版改为他引率)纳入期刊评价指标^[17],从而方便人们了解各期刊自引在总被引频次中所占比例。JCR的后台数据不但给出了总被引频次中的自引量,而且给出了影响因子构成中的自引量,新版InCite JCR中还发布了各期刊历年来他引影响因子(impact factor without journal self cites)。这些举措无疑都是为了揭示各期刊的自引水平,警示和阻止期刊的过度自引。

尽管如此,通过提高期刊自引人为操纵影响因子的行为却愈演愈烈,甚至有的期刊为了避免自引率过高,有意引导作者引用自己期刊前2年发表的论文做参考文献。这样,既能有效提高期刊影响因子,又最大限度地避免了期刊自引率的过度升高;但是,任何人为操作都必然会反映在不同的期刊计量学指标上。假设有人

通过提高期刊自引来提升影响因子,又要控制期刊的自引率,必然会出现影响因子构成中的自引率(self-citation rate for impact factor, $R_{SC,IF}$)明显高于总被引频次中的自引率(self-citation rate for total citations, $R_{SC,TC}$)。为了验证这一假设,我们对2015年版InCite JCR中收录的8778种自然科学期刊的 $R_{SC,IF}$ 和 $R_{SC,TC}$ 做了全方位比较和分析,并对相关的几个问题做了研究。

1 研究对象和方法

1.1 研究对象 2015年版InCite JCR中收录的所有自然科学期刊,包括SCIE的177个学科8778种。

1.2 研究方法

1.2.1 数据获取

1) JCR期刊评价指标。登录InCite数据库(<https://incites.thomsonreuters.com>),点击进入Journal Citation Reports,年份选择2015年,版本选择SCIE,定制指标(Customize Indicators)勾选全部指标,包括期刊名称、期刊缩写名称、ISSN、总被引频次(total cites, C_T)、影响因子、他引影响因子、5年影响因子、即年指标、论文数、被引半衰期、特征因子、论文影响分值、影响因子分区、平均影响因子百分位(average JIF Percentile, aJIFP),导出格式选择CSV(用默认程序Excel打开,并转换为Excel格式文件)。

2) JCR背景数据。登录ISI-JCR,点击相应期刊名称,把每种期刊的出版语言、所属国家和地区、所属学科(若属多个学科,该刊所有数据复制到每个学科)、

总被引频次中的自引量(self-cites of total cites, $C_{S,TC}$)、影响因子中的被引量(cites to years used in impact factor calculation, C_{IF})和影响因子的自引量(self-cites to years used in impact factor calculation, $C_{S,IF}$)等参数添加到上述Excel文档中。

1.2.2 $R_{SC,IF}$ 和 $R_{SC,TC}$ 的计算

$$R_{SC,IF} = C_{S,IF}/C_{IF}, R_{SC,TC} = C_{S,TC}/C_T$$

1.2.3 统计学方法 统计学软件为SPSS 18,数据分布的正态性检验采用Kolmogorov-Smirnov检验, $R_{SC,IF}$ 与 $R_{SC,TC}$ 差异比较采用非参数检验中的2个相关样本的秩检验(Wilcoxon检验),检验水准: $\alpha = 0.05$ 。

2 结果与分析

2.1 期刊被引量、自引量和自引率总体情况 2015年版InCites JCR共收期刊8778种,TC总和为4890万8974, $C_{S,TC}$ 总和为352万4685,集合 $R_{SC,TC}$ 为0.072, C_{IF} 为790万3091, $C_{S,IF}$ 为88万7421,集合 $R_{SC,IF}$ 为0.112; $R_{SC,TC}$ 总和为817.2,平均值为0.093(817.2/8778), $R_{SC,IF}$ 总和为1150.7,平均值为0.132,详见表1。无论集合值还是平均值,均为 $R_{SC,IF}$ 明显大于 $R_{SC,TC}$ 。 $C_{S,TC} = 0$ 的期刊有208种,多数为被引频次较低的期刊,但也有个别期刊被引频次较高,如美国的《Arthritis Rheum》, C_T 为45624, $C_{S,TC}$ 居然为0。 C_T 大于1万而 $C_{S,TC}$ 为0者还有2种期刊。 C_T 如此之高, $C_{S,TC}$ 为0令人费解。 $C_{S,TC} = 0$ 的期刊有458种,多数为 C_{IF} 较低的期刊,但有5种期刊 $C_{IF} > 1000$ 。

表1 8778种期刊被引量、自引量和自引率总体情况

指标	总和	最大值	最小值	中位数	平均值	标准差	变异系数
C_T	48 908 974	627 846	1	1 405	5 571.8	19 755.1	3.55
$C_{S,TC}$	3 524 685	63 718	0	85	401.5	1 737.0	4.33
C_{IF}	7 903 091	188 116	0	227	900.3	3 488.0	3.88
$C_{S,IF}$	887 421	17 320	0	21	101.1	443.3	4.38
$R_{SC,TC}$	817.2	1.000	0.000	0.066	0.093	0.092	0.99
$R_{SC,IF}$	1 150.7	1.000	0.000	0.097	0.131	0.123	0.93

2.2 期刊被引量、自引量和自引率分布 8778种期刊 C_T 、 $C_{S,TC}$ 、 C_{IF} 、 $C_{S,IF}$ 的分布情况见表2。 $C_T < 5000$ 的期刊有6834种(77.9%), $C_T \geq 1$ 万者有1046种(11.9%), $C_T \geq 10$ 万者仅有41种。 C_T 和 C_{IF} 最集中的频段都是100~999,而 $C_{S,TC}$ 和 $C_{S,IF}$ 最集中的频段是1~99。

8778种期刊的自引率分布见表3。 $R_{SC,TC} < 0.01$ 和 $R_{SC,IF} < 0.01$ 的期刊分别为607和614种,二者基本持平, $R_{SC,TC}$ 介于0.01~<0.05和0.05~<0.10的期刊明显多于 $R_{SC,IF}$ 介于0.01~<0.05和0.05~<0.10的期刊,而自 $R_{SC} 0.10 \sim < 0.20$ 区段开始, $R_{SC,IF}$ 介于

该区段的期刊明显多于 $R_{SC,TC}$ 相同区段的期刊,自引率越高,这一趋势越明显。 $R_{SC,TC} \geq 0.50$ 的期刊有59种,而 $R_{SC,IF} \geq 0.50$ 的期刊竟然有177种, $R_{SC,TC} \geq 0.20$ 的期刊有947种,而 $R_{SC,IF} \geq 0.20$ 的期刊达到1862种。理论上讲,同一种期刊不存在学科差异和其他引证环境的不同, $R_{SC,IF}$ 和 $R_{SC,TC}$ 应该相差不大,除非存在人为操纵。二者异常明显的差异,一定程度上说明SCI来源期刊的影响因子存在较明显的人为操纵倾向。

此外,8778种期刊中, $R_{SC,IF}$ 和 $R_{SC,TC}$ 应均大于0.20者844种。

表2 8778种期刊被引量和自引量的分布情况

指标	被引量和自引量区间							
	0	1~99	100~999	1000~4999	5000~9999	10000~49999	50000~99999	>100000
C_T	0	304	3325	3205	898	918	87	41
$C_{S,TC}$	208	4487	3370	623	63	25	2	0
C_{IF}	17	2538	4593	1382	156	86	5	1
$C_{S,IF}$	458	6686	1506	120	2	6	0	0

表3 8778种期刊 $R_{SC,TC}$ 和 $R_{SC,IF}$ 的分布情况

自引率	自引率区间					
	<0.01	0.01~<0.05	0.05~<0.10	0.10~<0.20	0.20~<0.50	≥0.50
$R_{SC,TC}$	607	2827	2485	1912	888	59
$R_{SC,IF}$	614	1710	2190	2385	1685	177

注:有17种期刊 C_{IF} 和 $C_{S,IF}$ 都等于0,因此未计算 $R_{SC,IF}$ 。

2.3 $R_{SC,IF}$ 和 $R_{SC,TC}$ 的比较 2015年版JCR-SCI收录的8778种期刊中, C_{IF} 和 $C_{S,IF}$ 都等于0的期刊有17种(无法计算 $R_{SC,IF}$), $R_{SC,IF} = R_{SC,TC}$ 者208种(2.4%), $R_{SC,IF} > R_{SC,TC}$ 者6954种(79.2%), $R_{SC,TC} > R_{SC,IF}$ 者1599种(18.2%), $R_{SC,IF}$ 和 $R_{SC,TC}$ 都等于0者203种。因 $R_{SC,IF}$ 和 $R_{SC,TC}$ 2组数据均不呈正态分布,故对其进行2个相关样本的Wilcoxon秩检验, $Z = 59.836, P < 0.001$,2组差异有统计学意义,说明SCI收录的8778种期刊 $R_{SC,IF}$ 显著大于 $R_{SC,TC}$ 。

通常情况下,办刊人为了提高影响因子而又最大限度地避免自引率过度增加,只能有意识地引导甚至强迫作者引用自己期刊前2年发表的文章作为参考文献,这必然导致 $R_{SC,IF}$ 过度增加。近80%的期刊 $R_{SC,IF} > R_{SC,TC}$,说明JCR收录的很多期刊存在人为操纵的可能。

2.4 $R_{SC,IF}$ 和 $R_{SC,TC}$ 的差异程度 C_T 较小(如小于50)的情况下,计算 $R_{SC,TC}$ 没有意义。同样, C_{IF} 较小的情况

下,计算 $R_{SC,IF}$ 也没有意义。如美国的《Acoustics》(ISSN 1541-0161)的 $C_{IF} = 2, C_{S,IF} = 1$,则 $R_{SC,IF} = 0.50$ (其实构成影响因子的自引量仅为1次,自引率竟达50%);因此,比较 $R_{SC,IF}$ 和 $R_{SC,TC}$ 的差异程度时,应该将 $C_{IF} < 50$ 的期刊剔除。8778种期刊中, $C_{IF} < 50$ 者1370种, $C_{IF} \geq 50$ 者7408种。 $R_{SC,IF}$ 大于 $R_{SC,TC}$ 者达6114种,二者差异最大的是《American Scientist》(美国科学家),其 $R_{SC,IF}$ 高于 $R_{SC,TC}$ 达13.3倍, $R_{SC,IF}$ 高于 $R_{SC,TC}$ 5倍以上的期刊共34种(见表4),其中Q1和Q2区期刊11种,Q3和Q4区期刊共23种,说明高影响力期刊一般不会通过提高 $R_{SC,IF}$ 对影响因子进行人为操纵。 $R_{SC,IF}$ 高于 $R_{SC,TC}$ 1倍以上者共1456种,占19.7(1456/7408)。当然,不能认为 $R_{SC,IF}$ 高于 $R_{SC,TC}$ 1倍以上就一定存在人为操纵的可能,但确实没有更加合理的理由来解释同一期刊 $R_{SC,IF}$ 和 $R_{SC,TC}$ 如此明显的差异。

表4 $R_{SC,IF}$ 和 $R_{SC,TC}$ 差异程度较大的期刊及其主要指标

排序	期刊缩写名	出版语言	所属国家	C_T	F_1	F_1 分区	R_{SC}	$R_{SC,IF}$	$R_D^{1)}$
1	AM SCI	英语	美国	1983	0.500	Q3	0.0040	0.0577	13.30
2	BIOMETRIKA	英语	英格兰	16241	1.130	Q3	0.0095	0.1264	12.33
3	J PHYS CHEM REF DATA	英语	美国	6295	3.290	Q2	0.0194	0.2157	10.13
4	J LABELLED COMPD RAD	多语种	英格兰	1578	1.532	Q4	0.0114	0.1235	9.83
5	J LUMIN	英语	荷兰	14731	2.693	Q1	0.0091	0.0984	9.82
6	J NEUROGASTROENTEROL	英语	韩国	641	1.771	Q3	0.0125	0.1340	9.73
7	Z NATURFORSCH C	英语	德国	2677	0.709	Q4	0.0056	0.0556	8.91
8	IEEE COMPUT GRAPH	英语	美国	1691	1.203	Q2	0.0201	0.1798	7.94
9	AM STAT	英语	美国	4099	1.215	Q2	0.0285	0.2278	6.98
10	ACTA CRYSTALLOGR A	英语	英格兰	13728	2.333	Q2	0.0207	0.1575	6.61
11	Z KRIST - NEW CRYST ST	英语	德国	598	0.122	Q4	0.0518	0.3818	6.37
12	EUR CYTOKINE NETW	英语	法国	1206	2.562	Q3	0.0033	0.0244	6.35
13	COMMUN ACM	英语	美国	12458	3.301	Q1	0.0076	0.0560	6.35
14	BRIT MED BULL	英语	英格兰	3727	2.921	Q1	0.0043	0.0315	6.34
15	COMPUTING	英语	奥地利	1472	0.872	Q3	0.0075	0.0543	6.26
16	CLIN INVEST MED	多语种	英格兰	911	1.191	Q4	0.0066	0.0472	6.16
17	AM CERAM SOC BULL	英语	美国	1145	0.789	Q3	0.0070	0.0500	6.16

表4(续)

排序	期刊缩写名	出版语言	所属国家	C_T	F_1	F_1 分区	R_{SC}	$R_{SC,IF}$	$R_D^{1)}$
18	CHEMOTHERAPY	英语	瑞士	1 378	0.992	Q4	0.015 2	0.108 3	6.11
19	J FAM PRACTICE	英语	美国	2 032	0.658	Q3	0.010 8	0.076 9	6.10
20	J AM SOC HORTIC SCI	英语	美国	5 083	1.000	Q2	0.025 2	0.177 8	6.06
21	DYNAM ATMOS OCEANS	多语种	荷兰	830	1.439	Q3	0.012 0	0.084 7	6.03
22	J PEDIATR ORTHOP B	英语	美国	957	0.449	Q4	0.015 7	0.108 9	5.95
23	J HORTIC SCI BIOTECH	英语	英格兰	2 103	0.458	Q3	0.028 5	0.193 9	5.80
24	STAT APPL GENET MOL	英语	美国	1 202	1.265	Q4	0.014 1	0.095 2	5.73
25	EVOL BIOINFORM	英语	新西兰	1 272	1.404	Q4	0.009 4	0.062 5	5.63
26	SIAM REV	英语	美国	6 151	3.488	Q1	0.004 2	0.028 0	5.62
27	ANN CLIN PSYCHIATRY	英语	美国	934	1.583	Q3	0.003 2	0.021 1	5.55
28	Q APPL MATH	英语	美国	2 071	0.654	Q3	0.012 1	0.078 4	5.50
29	J HERED	英语	美国	6 013	2.075	Q3	0.023 4	0.152 4	5.50
30	CAN J CHEM	多语种	加拿大	8 193	1.003	Q3	0.012 8	0.079 5	5.20
31	ADV AGRON	英语	美国	3 862	4.381	Q1	0.004 1	0.025 4	5.12
32	BRAIN RES BULL	英语	美国	8 546	2.572	Q3	0.014 0	0.085 5	5.09
33	BIOMETRICS	多语种	美国	17 456	1.360	Q3	0.016 6	0.101 0	5.08
34	PHYS TODAY	英语	美国	3 829	3.234	Q1	0.007 3	0.044 2	5.04

注:1) R_D 为差异程度,即 $R_{SC,IF}$ 高出 $R_{SC,TC}$ 的倍数,计算方法: $(R_{SC,IF} - R_{SC,TC}) / R_{SC,TC}$ 。

2.5 $R_{SC,IF}$ 与 $R_{SC,TC}$ 差异较大的期刊的国家和地区分布 在 $R_{SC,IF}$ 高于 $R_{SC,TC}$ 5 倍以上的 34 种期刊中,美国期刊 18 种,英格兰 6 种,德国和荷兰各 2 种,奥地利、加拿大、法国、韩国、瑞士和新西兰各 1 种,几乎都是发达国家。 $R_{SC,IF}$ 高于 $R_{SC,TC}$ 1 倍以上的期刊共 1 456 种,剔除 $R_{SC,IF} < 10\%$ 的期刊(无论 $R_{SC,IF}$ 与 $R_{SC,TC}$ 悬殊多大, $R_{SC,IF} < 10\%$ 并不过分),剩余 738 种,分布在 48 个国家和地区,超过 5 种期刊的国家和地区有 24 个,详

见表 5。可以看出: $R_{SC,IF}$ 高于 $R_{SC,TC}$ 1 倍以上的期刊总数美国、英国和荷兰最多,其次是日本和德国;但占各自国家 JCR 收录期刊总数百分比最高的是捷克、印度、日本、匈牙利和加拿大。我国和法国在这 24 个国家中, $R_{SC,IF}$ 高于 $R_{SC,TC}$ 1 倍以上的期刊占比最低。这说明,我国 SCI 来源期刊通过提高自引人为操纵影响因子的现象最不明显,亚洲国家中日本和印度存在明显的人为操纵期刊影响因子的倾向。

表 5 $R_{SC,IF}$ 高于 $R_{SC,TC}$ 1 倍以上期刊 ≥ 5 个的国家和地区

国家	期刊数 ¹⁾	JCR 期刊数 ²⁾	百分比/%	国家	期刊数	JCR 期刊数	百分比/%
捷克	9	40	22.5	巴西	11	110	10.0
印度	19	101	18.8	美国	235	2 931	8.0
日本	44	234	18.8	瑞士	16	211	7.6
匈牙利	6	32	18.8	韩国	7	96	7.3
加拿大	15	86	17.4	德国	43	591	7.3
俄罗斯	24	146	16.4	丹麦	5	69	7.2
新加坡	8	52	15.4	西班牙	5	72	6.9
奥地利	5	34	14.7	英格兰	126	1 817	6.9
新西兰	5	35	14.3	荷兰	47	692	6.8
意大利	14	120	11.7	澳大利亚	6	99	6.1
罗马尼亚	5	43	11.6	中国	11	185	5.9
波兰	14	135	10.4	法国	10	172	5.8

注:1) $R_{SC,IF}$ 高于 $R_{SC,TC}$ 1 倍以上的期刊数;2) JCR 中收录的该国家期刊总数。

3 结论

1) SCI 来源期刊中 $R_{SC,IF}$ 显著高于 $R_{SC,TC}$ 。无论是集合 $R_{SC,IF}$ 与集合 $R_{SC,TC}$ 比较,平均 $R_{SC,IF}$ 与平均 $R_{SC,TC}$ 比较, $R_{SC,IF}$ 与 $R_{SC,TC}$ 区段的期刊分布,还是 8 778 种期刊 $R_{SC,IF}$ 与 $R_{SC,TC}$ 的 Wilcoxon 检验均证实,SCI 来源期刊中 $R_{SC,IF}$ 显著高于 $R_{SC,TC}$ 。8 778 种期刊中, $R_{SC,IF}$ 大

于 $R_{SC,TC}$ 者达 6 954 种(79.2%), $R_{SC,IF}$ 超过 $R_{SC,TC}$ 1 倍以上者 1 456 种期刊,剔除 $R_{SC,IF} < 10\%$ 者依然有 738 种期刊。该研究从多个方面对 JCR-SCI 全部来源期刊的 $R_{SC,IF}$ 与 $R_{SC,TC}$ 进行了比较,证实了近 80% 的 SCI 来源期刊, $R_{SC,IF}$ 大于 $R_{SC,TC}$ 。

2) SCI 来源期刊存在明显的人为操纵影响因子的现象。众所周知,不同学科期刊由于学科规模大小不

同、发展阶段不同、科研人员引证习惯不同,导致不同学科期刊总被引频次和影响因子存在明显差异;但同一学科内期刊的各文献计量学指标具有可比性。对于同一期刊而言,其 $R_{SC,IF}$ 和 $R_{SC,TC}$ 应该大致相当,没有理由出现较大差异,除非有人为操纵因素存在。该研究结果提示,2015年JCR-SCI收录的8778种期刊中,出现超大范围的 $R_{SC,IF} > R_{SC,TC}$ 的现象,甚至个别期刊 $R_{SC,IF}$ 超出 $R_{SC,TC}$ 10以上是极不正常的。由此认为,SCI来源期刊中存在较为普遍的人为操纵影响因子的可能。

提高影响因子最直接、最简单的途径是提高期刊的自引量。根据影响因子的定义,计算影响因子的引证时间窗口是2年^[18-19],超过2年的文献被引用对提高影响因子是无效的;因此,人为操纵主要表现在引导甚至强迫作者引用目标期刊最近2年(统计当年除外)发表的文献。这样一来,必然导致 $R_{SC,IF}$ 明显大于 $R_{SC,TC}$ 。如果某期刊 $R_{SC,IF}$ 大于 $R_{SC,TC}$ 异常明显,则在很大程度上存在人为操纵因素。

事实上,很多情况下 $R_{SC,IF}$ 和 $R_{SC,TC}$ 差异不明显,但二者若都很高(如 $>20\%$),依然值得怀疑。该研究8778种期刊中, $R_{SC,IF}$ 和 $R_{SC,TC}$ 均大于 20% 者844种。一般认为,期刊自引率应 $<20\%$ ^[20], $R_{SC,IF}$ 和 $R_{SC,TC}$ 均大 20% 也是很异常的。 $R_{SC,IF}$ 或 $R_{SC,TC}$ 异常高,或者 $R_{SC,IF}$ 显著高于 $R_{SC,TC}$ 均存在人为操作的可能。

3)我国SCI来源期刊不存在明显的人为操纵影响因子的现象。本文研究结果提示:在 $R_{SC,IF}$ 超过 $R_{SC,TC}$ 1倍以上期刊数较多(≥ 5 个)的24个国家中,我国和法国期刊占SCI来源期刊百分比最低(5.9%和5.8%);因此认为,国内SCI来源期刊并不存在明显的人为操纵现象。人为操纵影响因子倾向较为明显的有捷克、印度、日本、匈牙利和加拿大等国家。

我国期刊存在过度自引^[21]和小集团内部期刊互引^[22],后者甚至成为期刊界公开的秘密;但是,我国SCI来源期刊人为操纵现象并不明显,可能是因为入选SCI很不容易,都异常珍惜SCI期刊身份,不希望因为人为操纵而被剔除。国内消化病学领域的某刊(英文版),曾经因为 R_{SC} 过高一度被JCR删除(2008年恢复),这对我国SCI期刊起到了较好的警示作用。

4 参考文献

- [1] SERVAES J. On impact factors and research assessment: at the start of volume 31 of telematics and informatics [J]. Telematics and Informatics, 2014, 31(1): 1
- [2] CAMPANARIO J M, Coslado M A. Benford's law and citations, articles and impact factors of scientific journals [J]. Scientometrics, 2011, 88(2): 421
- [3] BRODY S. Impact factor: Imperfect but not yet replaceable [J]. Scientometrics, 2013, 96(1): 255
- [4] BRADSHAW C J A, BROOK B W. How to rank journals [J]. PLoS ONE, 2016, 11(3): e0149852
- [5] EPSTEIN D. Impact factor manipulation [J]. Journal of the European Medical Writers Association, 2007, 16(3): 133
- [6] WALLNER C. Ban impact factor manipulation [J]. Science, 2009, 323(5913): 461
- [7] YANG D H, LI X, SUN X X, et al. Detecting impact factor manipulation with data mining techniques [J]. Scientometrics, 2016, 109(3): 1989
- [8] Falagas ME, Alexiou VG. The top-ten in journal impact factor manipulation [J]. Archivum Immunologiae et Therapiae Experimentalis, 2008, 56(4): 223
- [9] YU G, WANG L. The self-cited rate of scientific journals and the manipulation of their impact factors [J]. Scientometrics, 2007, 73(3): 321
- [10] BARTNECK C, KOKKELMANS S. Detecting h-index manipulation through self-citation analysis [J]. Scientometrics, 2011, 87(1): 85
- [11] 王崇德. 科技文献的自引 [J]. 情报学刊, 1984, 5(1): 15
- [12] 罗式胜. 自引类型与分析 [J]. 情报科学, 1984, 5(3): 7
- [13] Yang D H, Li X, Sun X X. Detecting impact factor manipulation with data mining techniques [J]. Scientometrics, 2016, 109(3): 1989
- [14] Teodorescu D, Andrei T. An examination of "citation circles" for social sciences journals in Eastern European countries [J]. Scientometrics, 2014, 99(2): 209
- [15] 徐海丽. 影响因子人为操纵案例分析及构建期刊综合评价体系设想 [J]. 中国科技期刊研究, 2014, 25(5): 691
- [16] 刘雪立. 10种国际权威科技期刊影响因子构成特征及其启示 [J]. 编辑学报, 2014, 26(3): 296
- [17] 中国科学技术信息研究所. 1999年版中国科技期刊引证报告 [M]. 北京: 中国科技信息研究所, 1999
- [18] 刘雪立, 盖双双, 张诗乐, 等. 不同引证时间窗口影响因子的比较研究: 以SCI数据库眼科学期刊为例 [J]. 中国科技期刊研究, 2014, 25(12): 1509
- [19] DORTA-GONZÁLEZ P, DORTA-GONZÁLEZ M I. Impact maturity times and citation time windows: the 2-year maximum journal impact factor [J]. Journal of Informetrics, 2013, 7(3): 593
- [20] 金铁成. 此自引率非彼自引率: 学术期刊的自引率与被引率错引举隅 [J]. 编辑学报, 2016, 28(3): 251
- [21] 刘雪立, 周志新, 方红玲, 等. 2005—2007年我国医学期刊自引率与过度自引的界定 [J]. 中国科技期刊研究, 2009, 20(4): 624
- [22] 马峥. 通过计量指标分析发现操纵期刊评价结果的行为 [J]. 编辑学报, 2016, 28(6): 608

(2017-05-04 收稿; 2017-07-15 修回)