

## 论期刊影响因子人为操纵的识别\*

刘雪立

新乡医学院期刊社,河南省科技期刊研究中心,453003,河南新乡

**摘要** 近年来,期刊影响因子的人为操纵受到学术界广泛关注,期刊过度自引成为人为操纵影响因子的重要手段。期刊自引率是识别影响因子人为操纵的敏感指标,但对小集团内部期刊互引和其他形式人为操纵的识别却无能为力。在该研究中,深度挖掘了自引率、扩散因子、被引半衰期、开放因子和互引指数在期刊影响因子人为操纵识别中的应用。

**关键词** 学术期刊;影响因子;人为操作;自引率;扩散因子;被引半衰期;开放因子;互引指数

**On the recognition of journal impact factor manipulation // LIU Xueli**

**Abstract** In recent years, the manipulation of the journal impact factor has been widely concerned by academia, and the excessive self-citation of journals has become an important means of manipulating the impact factor. The self-citation rate of the journal is a sensitive index to identify the manipulation of impact factor, but it is incapable to recognize the internal journals in a small group and other forms of manipulation. In this study, the applications of self-citation rate, spreading factor, cited half-life, open factor and mutual citation index in the recognition of manipulation of journal impact factors are deeply detected.

**Keywords** academic journal; impact factor; manipulation; self-citation rate; spreading factor; half-cited life; open factor; mutual citation index

**Author's address** Periodicals Publishing House/ Henan Research Center for Science Journals, Xinxiang Medical University, Xinxiang 453003, Xinxiang, Henan, China

**DOI:** 10.16811/j.cnki.1001-4314.2018.01.030

自影响因子广泛应用于学术期刊评价以来,对影响因子科学性和合理性的质疑从来就没有间断过,争议最大的莫过于影响因子的滥用和误用<sup>[1-3]</sup>,如用影响因子评价单篇论文<sup>[4]</sup>、评价研究者个人<sup>[5]</sup>,甚至应用于职务任命、资金分配和项目评审<sup>[6]</sup>。许多研究者认为,影响因子的设计存在明显缺陷,如引证时间窗口(2年)过短<sup>[7-9]</sup>、分子和分母纳入的文献类型不统一<sup>[10]</sup>、没有考虑期刊论文被引频次的偏态分布<sup>[8,11]</sup>、无法应用于跨学科评价<sup>[12-13]</sup>等。

尽管如此,影响因子依然被广泛应用于期刊评价和研究绩效评价<sup>[14]</sup>。正像匈牙利学者 Brody<sup>[15]</sup>所说的那样,影响因子并不完美,但依然无法替代。澳大利

亚学者 Bradshaw 等<sup>[16]</sup>也指出: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。

正因如此,许多学术期刊的主编和编辑热衷于人为操纵期刊的影响因子<sup>[17-18]</sup>。希腊学者 Falagas<sup>[19]</sup>总结了人为操纵影响因子的10大行为,国内学者徐海丽<sup>[20]</sup>、郑毅<sup>[21]</sup>、鞠秀芳等<sup>[22]</sup>也对人为操作期刊影响因子的现象进行了研究。强制作者引用某期刊、小集团(甚至组成期刊联盟)内部期刊互引成为学术期刊界的潜规则,抑或是公开的秘密。

因此,如何从期刊的文献计量学指标识别可能的人为操纵就显得尤为重要,让人为操纵影响因子的行为无可遁形,以进一步加强期刊内涵建设、规范办刊行为。马峰<sup>[23]</sup>近期撰文,明确提出了通过计量指标分析发现操纵期刊评价结果的行为。他主要论证了通过期刊自引、他引的统计分析,揭示期刊过度自引、集团互引现象。

目前,国内外研究主要集中在对人为操纵行为的揭示和理论分析、通过对期刊自引率的研究判断期刊的人为操纵等方面,很少涉及其他指标在识别期刊影响因子和被引频次人为操纵中的作用。本研究将对自引率、扩散因子、被引半衰期、开放因子、互引指数等文献计量学指标在识别期刊影响因子人为操纵行为中的作用进行详细分析。

## 1 自引率与影响因子人为操纵的识别

**1.1 自引率的分类** 期刊自引率是指期刊总被引频次中自引次数所占的百分比。根据自引率适用对象的不同,我们将自引率分为2类,一是总被引频次中自引所占的百分比(self-cited rate for total cites,  $R_{SC,TC}$ ),二是计算影响因子的分子中自引所占的百分比(self-cited rate for impact factor,  $R_{SC,IF}$ )<sup>[24]</sup>。学术界普遍关注的是  $R_{SC,TC}$ ,对  $R_{SC,IF}$  研究较少。

《中国科技期刊引证报告(核心版)》中的是  $R_{SC,TC}$ ,清华同方的《中国学术期刊影响因子年报》中给出了他引影响因子,未给出  $R_{SC,TC}$ 。根据影响因子和他引影响因子可以计算出  $R_{SC,IF}$ 。2015年新版

\* 国家社会科学基金项目“影响因子缺陷的多维度矫正与学术期刊跨学科评价研究”(15BTQ061)

《JCR》(InCites JCR)里新增的3个指标中就包括他引影响因子,这样就能很方便地计算自引和他引对期刊影响因子的贡献度以及 $R_{SC,IF}$ 。

**1.2 自引率在影响因子人为操纵识别中的应用** 影响因子是指某期刊前2年发表的所有文献在统计当年的被引频次与该刊前2年发表的可被引文献数量之比<sup>[25-26]</sup>。显然,提高影响因子最根本的方法是增加期刊前2年发表的论文在统计当年的被引频次。其他期刊是否引用自家期刊不易控制,期刊自引就成为提高影响因子最直接、最有效的手段;因为,在论文发表的全过程涉及3类行为主体,即主编(编辑)、审稿专家、作者,其中作者是弱势群体,主编(编辑)起决定性作用。主编(编辑)要求作者引用某期刊时,作者为了能够顺利在该刊发表论文,一般没有勇气也没有必要拒绝主编(编辑)提出的引用要求,尽管有时候可能仅仅是建议。过度的人为操纵必然导致期刊自引率过度升高。一般来讲,期刊自引率应小于20%<sup>[27]</sup>,异常增高一定存在人为操作的可能。

为了遏制期刊通过提高自引量人为操作影响因子的行为,美国汤森路透在1997年版的《期刊引证报告》(Journal Citation Reports, JCR)里就加入了期刊总被引频次中的自引率( $R_{SC,TC}$ )和影响因子构成中的自引率( $R_{SC,IF}$ )。中国科学技术信息研究所在1999年出版的《中国科技期刊引证报告(核心版)》中专门增加了 $R_{SC,TC}$ 指标。国内消化病学领域某刊(英文版),曾经因为 $R_{SC,IF}$ 和 $R_{SC,TC}$ 过高一度被《JCR》剔除(2008年恢复),眼科学领域某中文期刊也曾因为 $R_{SC,TC}$ 过高一度被《中国科技期刊引证报告(核心版)》剔除。为了最大限度地发挥自引对影响因子的贡献,还不至于使自引

率(实际上仅指 $R_{SC,TC}$ )明显增加,部分办刊人开始强制或引导作者仅引用该刊前2年发表的论文作为参考文献,也就是马峥<sup>[23]</sup>文中提到的“通过在影响因子时间窗口内外调节自引操纵指标的行为”。这种行为一定会导致 $R_{SC,TC}$ 小幅增加而 $R_{SC,IF}$ 过度增加。通常情况下,同一期刊的 $R_{SC,TC}$ 和 $R_{SC,IF}$ 应该大体相当,如果 $R_{SC,IF}$ 远远大于 $R_{SC,TC}$ ,则可高度怀疑该刊存在人为操作行为。

## 2 扩散因子与影响因子人为操纵的识别

扩散因子是指期刊当年每被引100次所涉及的期刊数,体现期刊学术影响的集中度,也表征期刊影响力波及的范围<sup>[28]</sup>。该指标是中国科学技术信息研究所《中国科技期刊引证报告》中的特有指标,主要应用于识别期刊集团或期刊联盟内部期刊之间的互引。其实,期刊自引率高的实质,不是自己引自己多了,而是其他期刊引用它的少了。只要大幅度提高他引量,即使有大量的自引也不至于导致自引率过高。因此,各期刊为了充分发挥期刊自引对影响因子提升的作用,而又不致于导致自引率过高,只有关系不错的期刊之间结成期刊集团,集团内部期刊互相引用。这必然导致期刊总被引频次过度集中于少数期刊,扩散因子必将明显下降。如果某刊扩散因子明显低于同学科其他期刊,则可高度怀疑该刊有人为操作的可能。

表1汇总了《中国科技期刊引证报告(核心版)》中我国各师范大学学报(自然科学版)的相关文献计量学指标(按扩散因子降序排列),除了影响因子,其他指标均具有识别人为操作的功能。表1中的最后6种师范大学学报扩散因子均在50以下,最后2种期刊低得更加明显。

表1 我国师范大学学报(自然科学版)相关文献计量学指标

学报名称	影响因子	他引率	扩散因子	被引半衰期	开放因子	互引指数
1 师范大学学报	0.149	0.92	75.72	7.4	45	0.44
2 师范大学学报	0.120	0.97	75.52	7.3	49	0.64
3 师大学报	0.207	0.97	71.69	7.2	71	1.40
4 师范大学学报	0.236	0.86	70.12	4.6	33	0.84
5 师范大学学报	0.191	0.98	66.00	7.7	67	1.88
6 师范大学学报	0.205	0.85	63.83	7.5	39	0.56
7 师范大学学报	0.234	0.93	63.43	6.8	54	1.12
8 师范大学学报	0.327	0.92	63.23	6.4	55	1.24
9 师范大学学报	0.316	0.80	62.39	6.7	30	0.20
10 师范大学学报	0.202	0.82	61.54	6.7	36	1.28
11 师范大学自然科学学报	0.404	0.87	61.03	4.7	36	0.80
12 师范大学学报	0.336	0.90	58.66	6.8	59	1.76
13 师范大学学报	0.155	0.83	57.58	5.7	51	1.24
14 师范大学学报	0.254	0.85	57.49	6.9	43	1.00

表1(续)

学报名称	影响因子	他引率	扩散因子	被引半衰期	开放因子	互引指数
15 师范大学学报	0.376	0.90	55.47	6.0	54	1.20
16 师范大学学报	0.180	0.74	53.28	6.3	19	0.80
17 师范大学学报	0.307	0.90	52.94	7.9	54	1.04
18 师范大学学报	0.311	0.79	52.16	6.1	31	0.64
19 师范大学学报	0.239	0.82	50.93	6.4	21	0.44
20 师范大学学报	0.371	0.96	50.21	7.8	73	1.68
21 师范大学学报	0.290	0.60	49.28	5.9	11	0.28
22 师大学报	0.468	0.79	49.12	7.7	37	0.28
23 师范大学学报	0.393	0.83	48.85	4.3	27	1.20
24 师范大学学报	0.415	0.62	40.94	4.2	12	0.28
25 师范大学学报	0.346	0.56	35.48	5.4	4	0.12
26 师范大学学报	0.422	0.74	30.19	4.5	14	1.76

注:互引指数来源于2016年出版的《中国学术期刊影响因子年报》,其他指标均来源于2016年版《中国科技期刊引证报告》。

### 3 被引半衰期与影响因子人为操纵的识别

被引半衰期是指某期刊统计当年总被引频次中,较新的一半是在距离现在多长一段时间(以年为单位)内发表的<sup>[28]</sup>。被引半衰期通常用来测度期刊老化速度,它在期刊影响因子人为操纵识别方面的作用几乎未被关注。前文中提到一种人为操作影响因子的行为,即为了充分发挥自引在提升期刊影响因子中的作用,而又不使期刊自引率明显升高,必须尽可能多地引用该刊前2年发表的论文。这样一来,期刊总被引频次中必然过多地出现引用该刊最近2年发表的文献,3年前的文献引用明显偏少,直接导致被引半衰期明显下降,实际上就是人为地加速期刊文献老化的速度;因此,被引半衰期若明显低于同学科其他期刊,很可能存在人为操作影响因子的行为。表1中有5种师范大学学报(第4、11、23、24和26)被引半衰期(<5.0)明显小于其他师范大学学报。

### 4 开放因子和互引指数与影响因子人为操纵的识别

开放因子是指期刊被引用次数的一半所分布的最小施引期刊数量,体现学术影响的集中度<sup>[28]</sup>。确定开放因子的方法是,将引用某期刊的期刊按照施引次数降序排列,累计施引次数到50%时涉及的期刊数。开放因子是《中国科技期刊引证报告》中的特有指标。

互引指数是某期刊的被引集中度与该期刊所属学科的平均被引集中度之比,用于分析某期刊被引期刊分布广度的合理性<sup>[29]</sup>。某期刊的被引集中度是指将引用某期刊的期刊按引用该刊的次数由大到小排列,累加该期刊被引次数至总被引频次的50%所对应的期刊数。某学科期刊平均被引集中度为该学科期刊被

引集中度的算数平均数。互引指数是《中国学术期刊影响因子年报》中的特有指标。

从以上概念可知,开放因子和期刊被引集中度含义完全相同,只是数据来源不同。开放因子和互引指数用于不同学科期刊评价均无意义,同一学科期刊评价中,期刊被引集中度和互引指数是等效的。

无论期刊的被引频次过度集中于1家期刊(期刊自引)还是过度集中于少数几家期刊(小集团互引),被引频次累计到总被引频次50%时涉及的期刊数均明显下降,导致开放因子和互引指数明显低于本学科其他期刊;因此,开放因子和互引指数明显低于本学科其他期刊,可高度怀疑有人为操纵的可能。表1中的第16、19、21、23~26等7种师范大学学报开放因子不足30,第9、21、22、24和25等5种师范大学学报互引指数不足0.3。表现最为明显的是26师范大学学报,其影响因子在被观察的26种学报中位列第2位,但他引率(倒数第4位)、扩散因子(倒数第1位)、被引半衰期(倒数第3位)和开放因子(倒数第4位)都较低,可以肯定该刊一定有大量自引和小集团互引。由于其互引指数(来源于清华同方的《中国学术期刊影响因子年报》)较高,可以进一步确定该刊与大量一般期刊而非统计源期刊的期刊互引。若是与大量的统计源期刊互引,其开放因子不应该那么低。

笔者认为,他引率、扩散因子、被引半衰期、开放因子和互引指数等5个指标中:若同时有3个及以上指标明显低于同学科其他期刊,则基本可以确定存在人为操纵行为;若2个指标明显低于同学科其他期刊,则可高度怀疑存在人为操纵。

总之,现有的文献计量学指标体系已比较完善,只要我们深刻理解各指标的实质内涵,深度挖掘不同指

标在期刊人为操纵行为识别中的作用,任何人为操纵期刊被引频次和影响因子的行为都将无处藏身;但是,如果“期刊互引集团”规模足够大,任何指标都将失去作用。不过,这种情况下,各期刊指标均有所提高,对每一家期刊也就失去了比较优势;此外,“期刊互引集团”过大很容易暴露。这就决定了不太可能出现规模过大的“期刊互引集团”。

## 5 参考文献

- [1] MARKS M S, MARSH M, SCHROER T A, et al. Misuse of journal impact factors in scientific assessment [J]. *Traffic*, 2013, 14(6):611
- [2] SHUBERT E. Use and misuse of the impact factor [J]. *Systematics and Biodiversity*, 2012, 10(4):391
- [3] MALAY D S. Impact factors and other measures of a journal's influence [J]. *Journal of Foot and Ankle Surgery*, 2013, 52(3):285
- [4] SMITH R. Commentary: the power of the unrelenting impact factor, is it a force for good or harm? [J]. *International Journal of Epidemiology*, 2006, 35(5):1129
- [5] HOLDEN G, ROSENBERG G, BARKER K, et al. Should decisions about your hiring, reappointment, tenure, or promotion use the impact factor score as a proxy indicator of the impact of your scholarship? [J]. *Medscape General Medicine*, 2006, 8(3):21
- [6] ADAM D. The counting house [J]. *Nature*, 2002, 415(6873):726
- [7] 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
- [8] VANCLAY J K. Impact factor: outdated artefact or stepping-stone to journal certification? [J]. *Scientometrics*, 2012, 92(2):211
- [9] 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
- [10] LIU X L, GAI S S, ZHOU J. Journal impact factor: do the numerator and denominator need correction? [J]. *PLoS One*, 2016, 11(3):e0151414
- [11] 刘雪立, 周晶, 盖双双, 等. 基于被引频次分布特征的影响因子缺陷矫正: 位置指标用于期刊评价的实证研究 [J]. *图书情报工作*, 2016, 60(9):99
- [12] PYO S, LEE W, LEE J. A novel journal evaluation metric that adjusts the impact factors across different subject categories [J]. *Industrial Engineering and Management Systems*, 2016, 15(1):99
- [13] LEYDESDORFF L, ZHOU P, BORNMANN L. How can journal impact factors be normalized across fields of science? An assessment in terms of percentile ranks and fractional counts [J]. *Journal of the American Society for Information Science and Technology*, 2013, 64(1):96
- [14] 刘雪立. 全球性 SCI 现象与影响因子崇拜 [J]. *中国科技期刊研究*, 2012, 23(2):185
- [15] BRODY S. Impact factor: imperfect but not yet replaceable [J]. *Scientometrics*, 2013, 96(1):255
- [16] BRADSHAW C J A, BROOK B W. How to rank journals [J]. *PLoS ONE*, 2016, 11(3):e0149852
- [17] OPTHOF T. Inflation of impact factors by journal self-citation in cardiovascular science [J]. *Netherland Heart Journal*, 2013, 21(4):163
- [18] YANG D H, LI X, SUN X X, et al. Detecting impact factor manipulation with data mining techniques [J]. *Scientometrics*, 2016, 109(3):1989
- [19] FALAGAS M E, ALEXIOU V G. The top-ten in journal impact factor manipulation [J]. *Archivum Immunologiae Et Therapiae Experimentalis*, 2008, 56(4):223
- [20] 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
- [21] 郑毅. 科技期刊人为提高自引率对影响因子的定量研究 [J]. *江汉大学学报(自然科学版)*, 2012, 40(4):52
- [22] 鞠秀芳, 郑彦宁, 潘云涛. 期刊引用操纵行为研究综述 [J]. *西南民族大学学报(人文社会科学版)*, 2013(4):224
- [23] 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
- [24] 刘雪立, 魏雅慧, 盛丽娜, 等. 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 [J]. *编辑学报*, 2017, 29(6):603
- [25] CAMPANARIO J M. Providing impact: the distribution of JCR journals according to references they contribute to the 2-year and 5-year journal impact factors [J]. *Journal of Informetrics*, 2015, 9(2):398
- [26] 刘雪立. 基于 Web of Science 数据库预测期刊影响因子的方法 [J]. *科技与出版*, 2014(2):87
- [27] 金铁成. 是自引证率, 还是自被引率?: 对加菲尔德的期刊自引率论断的考证 [J]. *中国科技期刊研究*, 2016, 27(7):704
- [28] 潘云涛, 马峥. 2016 年版中国科技期刊引证报告: 核心版 [M]. 北京: 科学技术文献出版社, 2016:7
- [29] 肖红. 研制说明 [J]. *中国学术期刊影响因子年报(自然科学与工程技术)*, 2016, 14:X

(2017-06-04 收稿; 2017-08-08 修回)