

基于核心论文数量和被引频次双重分布的百分位数指标改进研究

汪 莉

收稿日期:2017-08-25

修回日期:2018-03-12

南京师范大学图书馆,江苏省南京市栖霞区文苑路1号 210023

摘要 【目的】改进现有百分位数指标在揭示期刊载文量特征方面的不足,构建基于被引频次与载文量双重分布的期刊评价指标 PRC。【方法】以 2016 版《期刊引证报告》生物学学科的 85 种期刊为研究样本,通过二八定律筛选核心论文样本集,利用布拉德福区域分析法设计基于文献密度分布的载文量权重,利用百分位数建立基于被引频次分布的引文指标,构建期刊综合影响力评价指标 PRC。比较 PRC 与其他文献计量指标的相关性,验证指标评价效果。【结果】PRC 不仅在一定程度上克服了影响因子的内在缺陷,也能合理揭示期刊的出版规模效应,很好地均衡了相对值和绝对值两类指标,改进了百分位数指标的评价效果。【结论】PRC 指标兼顾被引频次和载文量双重分布,从数量和质量两个维度衡量期刊影响力,可用于评价科技期刊的学术影响力。

关键词 被引频次;载文量;核心论文;百分位数;PRC 指标

DOI: 10.11946/cjstp.201708250728

影响因子是衡量期刊学术影响力的重要指标之一,在国内外科研绩效评价中占据重要地位。然而随着科技进步与发展,影响因子越来越多地被滥用和误用,暴露出种种设计缺陷并饱受诟病^[1-3],人们正在不断寻求更合理的文献计量学指标。近年来,有学者提出用反映引文分布规律的百分位数替代影响因子的算术平均值^[4],受到文献计量学研究人员的广泛关注。百分位数是一种非参数统计方法,在个人、机构和学科领域中都得到了广泛运用^[5-8],包括美国国家自然科学基金会(National Science Foundation, NSF)^[9]、InCites 和基本科学指标(Essential Science Indicators, ESI)数据库^[10]等。德国文献计量学专家 Bornmann 等^[11]认为百分位数是实现学科领域、文献类型和出版年等引文信息标准化的最合理方法。目前,基于百分位数的期刊评价指标主要有百分位等级分数(Percentiles Rank Score, PRS)和综合影响指标(Integrated Impact Indicator, I3)。

PRS 将所有论文按照被引频次降序排列,计算单篇论文的百分位数并划分至不同的百分位等级区间。通过对各百分位等级的文献分布概率进行加权平均,可以得到各期刊的 PRS 数值。根据不同的研究目的,Bornmann^[12]提出以 6 个百分位等级为基础的 PR6 (Percentile Rank 6) 指标,荷兰情报学家

Leydesdorff 等^[13]建立了基于 100 个百分位等级的 PR100(Percentile Rank 100) 指标。PRS 可以定量描述文献被引频次在不同百分位位置的分布情况,在一定程度上解决了因被引频次偏态分布而造成的评价不公问题^[14]。但 PRS 在本质上仍是相对值指标,没有弥补影响因子在揭示研究对象体量和规模方面的不足。有学者因此质疑 PRS 的评价效果,如 Leydesdorff 等^[15]认为中心趋势度量指标对于数据集中趋势较为敏感,不适合用于描述偏态分布的数据。且所有引用被平均计数,对于出版量大的评价对象来说并不公平。为改进 PRS 指标,Leydesdorff 等^[15]进一步提出用“总量”定义期刊影响力的方案,并设计了 I3。I3 将所有论文划分为 100 个百分位等级,并分别赋予 1~100 的连续权值,但与 PRS 不同的是,计算公式中各百分位等级的文献相对分布概率被替换为论文绝对数量。I3 首次将论文数量纳入百分位数指标,评价结果同时兼顾到质量评价和数量评价,在一定程度上弥补了 PRS 的不足。当然,I3 设计也并非完美。作为绝对值指标,I3 对于论文的数量累积极为敏感,评价结果受到出版规模的直接影响,走向过于重视“多而全”的极端。

综上所述,虽然百分位数以分布来描述被引频次,较单纯的算术平均值有了长足进步,但仍存在一定局限性,相对值指标和绝对值指标的评价结果都

作者简介:汪 莉(ORCID:0000-0003-4941-2906),博士,研究馆员,E-mail:wangli@njnu.edu.cn。

容易受到文献集中载文量因素的影响。被引频次和载文量是科技期刊的两大基本要素,事实上合理的出版规模对于期刊影响力评价至关重要:期刊要实现学术影响力的质的飞跃,必须达到一定的体量和信息量;出版规模大又可能导致大比例零被引和低被引论文,抵消原有的评价优势,使评价指标偏离设计原意。设置合理的载文量指标,改进百分位数指标的不足,平衡相对值指标和绝对值指标的评价效果,是完善当前期刊评价体系的关键环节之一。

本研究拟借助传统文献计量学理论,筛选能够代表期刊真实影响力的核心论文,构建兼顾被引频次和载文量双重分布的期刊综合影响力评价指标。以 2016 版《期刊引证报告》(Journal Citation Reports, JCR)收录的 85 种生物学期刊为例,通过二八定律筛选各期刊内核心论文样本,用布拉德福区域分析法设计基于文献密度分布的载文量权重 C ,用百分位数指标建立基于被引频次分布的引文指标 PR6,从数量与质量两个维度构建新的期刊影响力评价指标 PRC (Percentile Rank C)。比较 PRC 与 5 年影响因子(5-Year Impact Factor, IF5)及其他百分位数指标的相关性,分析并验证新指标评价效果。

1 研究方法

从数量与质量两个维度来评价科技期刊影响力,是文献计量学发展的必然要求。本研究提议在兼顾引文分布的同时,赋予不同文献密度的期刊以不同权重,希望通过这一新变量凸显出版规模对于期刊综合影响力的贡献度。通过设置载文量权重考察期刊文献密度分布,可以从数量维度评价期刊;通过百分位数指标考察期刊引文分布情况,可以从质量维度评价期刊。两者的有机结合能够全面、科学地反映期刊整体影响力。

基于以上思路,设计分别代表期刊文献数量和质量的的评价指标——载文量权重 C 和百分位数指标 PR6,在此基础上建立新的期刊评价指标 PRC。

1.1 筛选核心论文样本集

期刊论文的引用分布反映出期刊发表后存在着核心论文的事实^[16]。事实上,期刊学术影响力是由少数高被引论文决定的,大量零被引论文与低被引论文不仅没有对期刊影响力产生贡献,反而会抵消原有的评价优势。因此,建立在核心论文基础上的期刊评价,不仅是期刊被引用情况和学术影响力的真实体现,也能在一定程度上减少因出版规模差异、

论文水平参差不齐而导致的评价失真。二八定律是文献计量学的重要定律之一,有研究表明二八定律能够有效区分不同被引频次论文的学术价值,且一定时期内相对稳定^[17],本研究以此作为筛选期刊核心论文的方法。

考虑到学术论文的引证规律,不同时间发表的论文,其被引峰值存在差异。如果不加区别地运用二八定律,必然导致近期发表的论文被遗漏。因此以期刊为单元,对所有论文进行历时法分析,逐年筛选出累积被引频次排序在前 20% 的核心论文,从而得到各期刊的核心论文。以 *Faseb Journal* 为例,该刊 2011—2015 年间发表文献总计 2337 篇,其中 2011 年发表论文 428 篇,取被引频次排序在前 20% 的 86 篇为核心论文,2012 年发表论文 494 篇,取被引频次排序在前 20% 的 99 篇为核心论文,其他年度以此类推,共得到 5 年周期的核心论文 467 篇,将其视为 *Faseb Journal* 的核心论文样本集。

1.2 载文量权重

布拉德福区域分析法将同一学科主题的期刊按照载文量降序排列,划分至论文数量相等的 3 个区域,即核心区、相关区和外围区,各区期刊的数量关系被总结为等比数列 $n_1:n_2:n_3=1:a:a^2$ (n 为各区期刊数, a 为布拉德福常数)。

英国文献学家 Vickery^[18] 进一步指出布拉德福期刊分区可以划分为任意数字,本研究据此将参与评价的所有期刊划分为 6 个区域,分别赋予相应的载文量权重(表 1)。具体做法如下。

(1) 将所有期刊按照其拥有的核心论文数量降序排序。

(2) 根据各分区核心论文数相等的原则,将期刊划分为 6 个区域,使各分区核心论文数恰好等于核心论文总数的 1/6。拥有核心论文数最多的 n_1 种期刊被划分至第 1 区,其他依次递减,拥有核心论文最少的 n_6 种期刊被划分至第 6 区。

(3) 第 1 区的期刊数量最少,核心论文密度最高,载文量权重取值为 6,其他以此类推,第 6 区的期刊数量最多,核心论文密度最低,载文量权重取值为 1。

1.3 被引频次指标 PR6

将所有核心论文按照被引频次降序排列。采用历时法,逐年计算样本期刊中单篇论文的百分位数,被引频次相同的文献,百分位数取平均值,最后将核心论文样本按照百分位数值降序排列。

表1 基于布拉德福区域分析法的期刊分区情况

核心论文数量	分区	期刊名称	分区期刊数量	载文量权重 C
各分区核心论文数量相等,即为核心论文总数的1/6	第1区	期刊 A1	n_1	6
		期刊 A2		
			
	第2区	期刊 B1	n_2	5
		期刊 B2		
		期刊 B3		
			
	第3区	期刊 C1	n_3	4
		期刊 C2		
		期刊 C3		
		期刊 C4		
			
	第4区	期刊 D1	n_4	3
		期刊 D2		
		期刊 D3		
		期刊 D4		
		期刊 D5		
			
	第5区	期刊 E1	n_5	2
		期刊 E2		
		期刊 E3		
		期刊 E4		
		期刊 E5		
		期刊 E6		
			
	第6区	期刊 F1	n_6	1
		期刊 F2		
		期刊 F3		
期刊 F4				
期刊 F5				
期刊 F6				
期刊 F7				
.....				

采用 NSF 的做法,将样本期刊划分为 6 个百分位等级,即 top1%、99%~95%、94%~90%、89%~75%、74%~50%和 bottom50%,分别代表核心论文集中被引频次分布在 top1%、top5%、top10%、top25%、top50%和 bottom50%的论文。计算公式为

$$\eta_{PR6} = \sum_{x=1}^k x \cdot f(x), \quad (1)$$

式中 η_{PR6} 为被引频次指标 PR6, x 为 $1-k$ 的离散随机变量,即 6 个百分位等级, f 为 x 出现的概率,当论文处于 top1%等级时, f 为 6,当论文处于 top5%等级时, f 为 5,其他以此类推。

1.4 PRC

PRC 将载文量权重与引文指标进行有机结合,计算公式为

$$\eta_{PRC} = \eta_{PR6} \cdot C, \quad (2)$$

式中 η_{PRC} 为 PRC 指标。由(2)式可知,PRC 由载文

量和被引频次两个因素共同决定,是一个数量与质量相结合的评价指标,数值越高表明期刊整体影响力越高,反之则表明期刊整体影响力越低。

2 实证研究

2.1 数据来源

学界普遍认为 5 年的时间跨度比较适合评价科技期刊,因此采用 5 年周期作为数据统计窗口。选取 2016 年版 JCR 的 85 种生物学期刊为研究对象。登录 Web of Science (WoS) 数据库,检索出 85 种样本期刊在 2011—2015 年间出版的所有论文,仅保留研究论文(Article)和综述论文(Review)两种类型的文献,得到 51250 篇“全样本”数据集。通过 WoS 平台的“创建引文报告”功能,将所有论文的逐年被引频次导出到 Excel,自动计算出 51250 篇论文自发表年至 2016 年底的累计被引频次。对全样本进行历时法分析,通过二八定律划分各期刊核心论文,得到 85 种期刊的核心论文样本共 10250 篇。

2.2 计算结果

根据(2)式对基于 85 种生物学期刊核心论文的计算结果进行实证研究,所得排名前 20 的期刊见表 2。为比较评价结果与其他评价指标的异同,同时计算了基于“全样本”数据的总载文量 N 、总被引频次 P 、PR6、PR100 和 I3 指标。其中 PR6、PR100 和 I3 指标的计算方法来源于文献[4,15]的研究成果。IF5 来源于 2016 版 JCR,统计 2011—2015 年间发表的论文在 2016 年底的期刊总被引频次,与本研究的数据统计时间窗口保持一致。

2.3 结果分析

PRC 排名前 3 的 *PLoS Biology*、*Elife* 和 *Proceedings of the Royal Society B-Biological Sciences* 都是公认的生物学顶级期刊,符合学界对学术期刊的主流看法,三者 IF5 排名分别为第 3、4、10。IF5 排名第 1 和第 2 的是综述类刊物 *Biological Reviews* 和 *Physics of Life Reviews*,由于综述类期刊载文量低,被引频次高,影响因子也往往偏高,但两者在 PRC 排名中分别降至第 8 和第 16,表明 PRC 对影响因子的设计缺陷有明显的矫正作用。

6 个指标中,总载文量 N 、总被引频次 P 和 I3 是具有累积规模优势的绝对值指标。总载文量和总被引频次排名第 1 的 *Proceedings of the Royal Society B-Biological Sciences*,其 I3 排名也是第 1。这是因为

I3 对期刊影响力的定义为“总和”,重视文献的数量累积。也正因为出版的文献总量多,导致 I3 排名靠前期刊的 IF5、PR6 和 PR100 是可以消除规模优势的平均值指标,三者排序第 1

的都是综述类期刊 *Biological Reviews*。总体说来,以上 20 种期刊中,有 14 种期刊的 PRC 排名介于相对值指标和绝对值指标之间,说明 PRC 对两类引文指标有较好的平衡作用。

表 2 PRC 计算结果及对比(以 PRC 值降序排列)

期刊名	总载文量 N 排序	总被引频 次 P 排序	PRC		IF5		PR6		PR100		I3	
			数值	排序	数值	排序	数值	排序	数值	排序	数值	排序
<i>PLoS Biology</i>	8	4	20.115	1	10.206	3	2.931	3	64.009	6	841.72	7
<i>Elife</i>	3	5	19.888	2	8.385	4	2.000	19	42.828	41	1218.02	4
<i>Proceedings of the Royal Society B-Biological Sciences</i>	1	1	17.378	3	5.417	10	2.785	5	70.800	4	2153.75	1
<i>Philosophical Transactions of the Royal Society B-Biological Sciences</i>	6	3	17.113	4	6.920	6	3.060	2	75.299	2	1435.20	5
<i>Faseb Journal</i>	4	2	14.807	5	5.435	9	2.824	4	71.910	3	1676.95	2
<i>Journal of Experimental Biology</i>	2	6	11.980	6	3.482	15	2.076	17	52.486	18	1584.04	3
<i>Bioessays</i>	18	9	11.411	7	4.597	11	2.329	10	57.611	15	469.53	10
<i>Biological Reviews</i>	54	10	9.700	8	11.740	1	3.579	1	82.201	1	245.78	30
<i>Biology Letters</i>	7	8	9.539	9	3.330	17	2.206	14	56.855	16	831.79	8
<i>BMC Biology</i>	32	11	8.838	10	7.556	5	2.442	8	61.953	9	344.46	16
<i>Journal of Theoretical Biology</i>	5	7	8.157	11	2.260	26	1.855	20	50.828	20	1130.42	6
<i>Chronobiology International</i>	25	14	6.458	12	3.056	20	2.155	15	58.594	13	384.96	13
<i>Bioscience</i>	40	12	6.457	13	6.574	7	2.609	6	63.869	7	292.52	21
<i>Computers in Biology and Medicine</i>	9	15	6.055	14	1.953	33	1.656	37	43.688	39	477.95	9
<i>Science China-Life Sciences</i>	17	16	5.964	15	2.014	29	1.759	27	48.753	25	402.70	12
<i>Physics of Life Reviews</i>	33	35	5.802	16	10.617	2	1.419	60	35.232	61	186.73	42
<i>Radiation Research</i>	19	13	5.748	17	2.775	21	2.126	16	57.778	14	449.51	11
<i>Bioelectrochemistry</i>	44	17	4.747	18	3.413	16	2.343	9	63.010	8	258.97	28
<i>Brazilian Journal of Medical and Biological Research</i>	13	24	4.655	19	1.469	47	1.557	49	40.898	49	363.99	14
<i>Microscopy Research and Technique</i>	15	25	4.628	20	1.235	58	1.559	48	41.635	46	358.48	15

为进一步比较 PRC 指标的评价效果,选取 3 种代表性期刊进行详细分析,分别为 *PLoS Biology* (PRC 排名第 1)、*Proceedings of the Royal Society B-Biological Sciences* (总载文量 N 、总被引频次 P 和 I3 排名均为第 1) 和 *Biological Reviews* (IF5、PR6 和

PR100 排名第 1)。图 1 所示为 3 种期刊不同被引频次层次论文的数量分布情况,如果某期刊在 top1%~10%等级的论文分布越多,在 top25% 及以后等级的论文分布越少,则表示该刊发表的高水平论文越多,整体影响力越高。

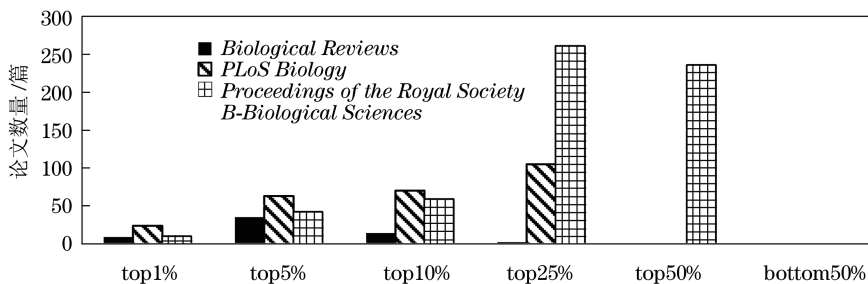


图 1 3 种期刊在不同研究水平上的论文数量分布

由图 1 可知,3 种期刊中 *Biological Reviews* 整体载文量最少,总被引频次最低,但所有论文全部分布在 top1%~10% 等级,反映了综述类期刊载文量少、篇均被引频次高的一般性特点。*Proceedings of the Royal Society B-Biological Sciences* 期刊论文规模最大,总被引频次最高,但绝大多数论文分布在

top25% 及 top50% 等级,表明该刊低被引论文所占比例高,整体研究绩效较低。相较而言,PRC 排序第 1 的 *PLoS Biology* 不仅具有一定的累积优势,而且研究绩效最高,表现为在 top1%~10% 等级高水平论文数量最多,在 top25% 及以后等级低水平论文数量少。由此可见,学术期刊的出版规模要在保证论文

质量的同时,达到合理的体量规模和信息量,才能提升期刊整体影响力。*PLoS Biology* 是3种期刊中整体影响力最高的期刊。这一结果验证了PRC指标构建方式的合理性,表明PRC同时考虑到期刊论文的载文量与被引频次分布,比现有百分位数指标更

科学。

2.4 PRC与其他评价指标相关性分析

使用SPSS软件进行数据统计,采用Spearman秩相关性检验来分析PRC指标与总载文量 N 、总被引频次 P 、IF5及各百分位数指标的相关性,结果见表3。

表3 各指标Spearman相关系数($N=85$)

Spearman	总载文量 N	总被引频次 P	IF5	PR6	PR100	I3	PRC
总载文量 N	—	0.809**	0.319**	0.381**	0.355**	0.932**	0.895**
总被引频次 P	0.809**	—	0.729**	0.812**	0.776**	0.952**	0.924**
IF5	0.319**	0.729**	—	0.873**	0.815**	0.521**	0.647**
PR6	0.381**	0.812**	0.873**	—	0.974**	0.650**	0.660**
PR100	0.355**	0.776**	0.815**	0.974**	—	0.628**	0.626**
I3	0.932**	0.952**	0.541**	0.650**	0.628**	—	0.887**
PRC	0.895**	0.924**	0.647**	0.660**	0.626**	0.887**	—

注:**在置信度(双侧)为0.01时,相关性具有统计学意义。

Spearman相关性分析显示各指标间相关性具有统计学意义($P<0.001$)。经检验发现,PRC不仅与总载文量、总被引频次及I3等绝对值指标相关性很强,而且与IF5、PR6、PR100等相对值指标也呈正相关,相关系数均大于0.6。其他指标中,相对值指标之间或绝对值指标之间都具有较高的相关性,高于相对值指标与绝对值指标之间的相关性,其中载文量与所有相对值指标都不相关,I3与IF5不相关。值得注意的是,PRC是所有指标中唯一一个与其他指标都呈正相关的评价指标,相关系数介于0.626至0.924之间,表明PRC不仅与现有期刊评价指标具有较强相关性,而且也具有一定独立性。相关性分析结果说明,PRC在相对值和绝对值两类评价指标中取得了较好的平衡作用,能够同时兼顾论文数量和论文质量两大要素,是可用于期刊评价的优秀指标。

3 结论

更深层次的载文量信息的探讨对于当前学术期刊评价具有重要意义。20世纪30年代,布拉德福提出文献的集中与离散规律,奠定了以论文数量为中心的核心期刊理论基础。20世纪70年代以后,随着引文分析法在科研绩效评价领域的应用,期刊评价开始由数量评价过渡到质量评价。此后,基于被引频次的各类衍生指标大量出现,而关于载文量的讨论则日渐萎缩,相关研究一直停留在绝对论文数量的比较上,没有逃离布拉德福时代核心期刊理论的窠臼。

本研究从改进现有引文指标角度出发,借用传统文献计量学相关理论,设计了基于期刊核心论文分布的数量指标,通过这一变量来彰显载文量对于

期刊整体影响力的贡献。将载文量指标与引文指标有机结合,从数量与质量两个维度描述期刊影响力,不仅完善了当前基于百分位数指标的期刊评价体系,也从不同角度深化了对于载文信息的探讨。

理论分析和实证研究均表明,PRC兼顾了被引频次和载文量双重分布,不仅描述了被引频次在期刊文献中的偏态分布,也能合理揭示期刊的出版规模效应,在一定程度上克服了传统期刊评价指标——IF5的内在缺陷。针对百分位数指标的不足,PRC也显示出其优越性,很好地均衡了相对值指标和绝对值指标的评价效果,改进百分位数指标评价效果,为完善期刊整体影响力评价体系提供了新的思路。PRC也与IF5及百分位数指标呈现出良好的相关性,表明PRC是可以用于科技期刊学术影响力评价的新指标。

当然PRC指标仍有需要细化和完善之处。首先,PRC中体现被引频次分布的引文指标建立在NSF的百分位等级划分基础上。但本研究采用二八定律筛选核心论文样本集,已经剔除了期刊的零被引论文和部分低被引论文,文献数量减少导致百分位数前移,仍按照top1%、99%~95%、94%~90%、89%~75%、74%~50%和bottom50%这6个等级对文献进行划分,则top1%等级对于大多数期刊来说可能会过于苛刻,等级划分仍需进一步研究。其次,PRC设置6个布拉德福分区,并以1~6的连续变量作为布拉德福分区的载文量权重,这一做法更多是出于期刊评价实践中的经验性观察,其合理性仍需与同行评议结果进行对比验证。

参考文献

- [1] 李晓红,于善清. 慎重使用“影响因子”评价科技期刊[J]. 情

- 报科学,2005,23(1):75-77.
- [2] 史庆华. 影响因子评价专业学术期刊的科学性与局限性[J]. 现代情报,2006,26(1):35-36.
- [3] 刘雪立. 全球性 SCI 现象和影响因子崇拜[J]. 中国科技期刊研究,2012,23(2):185-190.
- [4] Bornmann L, Mutz R. Further steps towards an ideal method of measuring citation performance: the avoidance of citation (ratio) averages in field-normalization [J]. *Journal of Informetrics*, 2011,5(1):228-230.
- [5] 郑芳,周群,陈仕吉,等. 基于百分位数指标的个人科研绩效评价研究[J]. 科研管理,2013,34(S1):189-194.
- [6] 周群,左文革,陈仕吉. 基于百分位数的文献计量指标研究综述[J]. 现代图书情报技术,2013(Z1):82-88.
- [7] 舒予,张黎俐. 基于动态权重的百分位数指标在学术影响力评价中的应用[J]. 图书情报工作,2016,60(11):93-99.
- [8] 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——位置指标应用于期刊评价的实证研究[J]. 图书情报工作,2016,60(9):99-105.
- [9] Science and Engineering Indicators 2016[EB/OL]. [2018-01-14]. <https://www.nsf.gov/nsb/sei/infographic1/index.html>.
- [10] Clarivate. Percentiles[EB/OL]. [2018-01-14]. <https://esi.incites.thomsonreuters.com/BaselineAction.action>.
- [11] Bornmann L, de Moya Anegón F, Leydesdorff L. The new excellence indicator in the world report of the SCImago institutions rankings 2011 [J]. *Journal of Informetrics*, 2012, 6(2):333-335.
- [12] Bornmann L. Towards an ideal method of measuring research performance: some comments to the Opthof and Leydesdorff (2010) Paper [J]. *Journal of Informetrics*, 2010, 4(3):441-443.
- [13] Leydesdorff L, Bornmann L, Mutz R, et al. Turning the tables on citation analysis one more time: principles for comparing sets of documents[J]. *Journal of the Association for Information Science & Technology*, 2011, 62(7):1370-1381.
- [14] Leydesdorff L, Bornmann L. Percentile ranks and the integrated impact indicator (I3) [J]. *Journal of the Association for Information Science & Technology*, 2012, 63(9):1901-1902.
- [15] Leydesdorff L, Bornmann L. Integrated impact indicators compared with impact factors: an alternative research design with policy implications[J]. *Journal of the Association for Information Science & Technology*, 2011, 62(11):2133-2146.
- [16] 董建军. 科技期刊文献引用分布规律的探讨[J]. 中国科技期刊研究,2013,24(4):688-693.
- [17] 杨利军,万小渝. 期刊论文低被引标准的界定方法[J]. 情报理论与实践,2013,36(7):51-53.
- [18] Vickery B C. Bradford's law of scattering [J]. *Journal of Documentation*, 1948, 4(3):198-203.

Improvement of percentile indicators based on double-distribution of number of core papers and citation frequencies

WANG Li

Library of Nanjing Normal University, 1 Wenyuan Road, Qixia District, Nanjing 210023, China

Abstract: [Purposes] This paper aims to improve the defects of percentile indicators in revealing the characteristic of number of published papers, and introduce a new indicator, Percentile Rank C (PRC), which is based on double-distributions of citation frequencies and number of published papers. [Methods] We chose 85 biology journals indexed in JCR (2016 edition) as samples. The core papers in these journals were selected by the Pareto principle. The weight of number of published articles based on the literature density distribution was designed by Bradford's law, and the indexes based on the citation frequency distribution were established by Percentile Rank 6. In aim to discuss the effects of PRC, we analyzed the correlations among PRC and other journal evaluation indicators. [Findings] The PRC indicator overcomes the intrinsic defects of impact factor, reasonably reveals the publication scale of journals, and fairly balances the evaluation effects between the relative value indicator and the absolute value indicator with an effect that improves the percentile indicators. [Conclusions] The PRC indicator takes the distribution of both number of published papers and citation frequencies into account, and journal influence can be measured by two dimensions of quantity and quality. It is a new indicator that can be used to evaluate the academic influence of scientific journals.

Keywords: Citation frequency; Number of published paper; Core paper; Percentile Indicator; Percentile Rank C

(本文责编:梁永霞)