

基于 F1000 与 WoS 的同行评议与文献计量相关性研究*

宋丽萍 王建芳

摘 要 为比较同行评议与文献计量方法在科学评价中的有效性及相关性,选取 F1000 以及 Web of Science 数据库,采用 SPSS16.0 软件,将近 2000 篇论文的 F1000 因子与 Web of Science 数据库中指标进行相关性比较。结果显示,F1000 因子与统计区间内的被引频次呈显著正相关,同时一些 F1000 因子很高的论文并没有高频被引,反之亦然。结论指出:从统计学的视角,文献计量指标与同行评议结果具有正向相关性,但是无论是同行评议还是文献计量,单独作为科学评价标准都会有失偏颇,以引文分析为代表的定量指标与同行评议方法的结合将是未来科学评价的主流。表 4。参考文献 14。

关键词 同行评议 文献计量 F1000 Web of Science

分类号 G250

The Correlation between Peer Review and Bibliometric Analysis in Evaluating Scientific Publication Outputs: A Case Study of F1000 and WoS

Song Liping & Wang Jianfang

ABSTRACT To compare expert assessment with bibliometric indicators, this paper selected a sample of about 2000 papers in F1000, and compared the initial assessment of the F1000 Article Factor by expert assessment to other measurements of the papers' impact using SPSS16.0. We found that there is a positive correlation between a paper's F1000 rating and its impact measured by number of citations. Despite the overall significantly positive correlations between assessments of importance and citations, the analysis shows that there are exceptions at the individual paper level: papers that were highly rated by expert reviewers were not always the most cited, and vice versa. These exceptions suggest that bibliometric measures by themselves may not be sufficient enough to measure research quality and importance, and tools that link expert reviews and more quantitative indicators, such as citation analysis, would be valuable additions to the field of research assessment and evaluation. 4 tabs. 14 refs.

KEY WORDS Peer review. Bibliometrics. F1000. Web of Science.

2006 年,英国高等教育资助委员会(HEFCE)提出:以同行评议为基础、具有 24 年历史的英国高等教育评估系统 RAE(Research Assessment Exercise)将被 REF(Research Excellence

Framework)所取代,届时将主要依据科研成果的引用程度,而非同行评审过程评估英国高校和研究机构的科研质量^[1]。以此为催化剂,并伴随着文献计量“寒武纪大爆炸”的到来及其在科

* 本文系天津市哲学社会科学规划课题“h 指数视角下的新一代科学评价指标与体系研究”(项目编号: TJTQ10-673)的研究成果之一。

通讯作者:宋丽萍,Email:slp-lp@163.com

研评价中份额的与日俱增^[2],引发了对科研评价中两种主要方法有效性及其关系的热烈探讨。2010 年, *the Scientist* 以及 *Nature* 杂志曾分设专题进行理论探讨。本文将采用 F1000 与 Web of Science 数据库,对文献计量与同行评议的相关性与一致性进行定量分析。

1 相关文献回顾及背景

关于同行评议与引文分析之间的关系,目前并无定论。文献[3]通过 RAE 与被引频次的分析,得出专家评议与被引量高度相关的结论,此成果不仅使 Oppenheim 成为 RAE 与引文分析正向相关的发现者,并使其因此萌生了以文献计量替代 RAE 的想法。文献[4]与上述结论相同。文献[5]则通过 2008 年英国 RAE 评议结果与 h 指数、g 指数的相关分析,指出二者相关性因学科而异。然而,文献[6]依据 *Physics in Medicine and Biology* 上刊载的 2003、2004 与 2005 年文献,分析指出:在 0.001 的显著性水平上,引用与专家评分负相关。文献[7]中的调查亦证实,在澳大利亚公共卫生 6 个分支中,其中两个领域同行评议与引文分析负相关或者不相关。至于 F1000 的应用,文献[8]通过 1530 篇文献,采用百分比分析,得出 F1000 至少在鉴定生态学文献方面是有局限性的,并将其归因于不同生态主题覆盖面的不均衡,任人唯亲,以及地域偏见,即倾向于北美出版物。上述研究表明,二者之关系已成为关注的焦点。

本文采用的数据库之一——F1000,全称 Faculty of 1000,又名千名科学家,来源于 Science Navigation Group 机构主席 Vitek Tracz 建立一个虚拟“机构”的设想,由世界上最好的 1000 名科学家在其中“任职”,主旨在于指引人们检索、发现该专业领域内有重要价值和前沿性的文献^[9]。据此,英国现代生物出版集团(BioMed Central)于 2002、2006 年分别推出“Faculty of 1000 Biology”(千名生物学家)和“Faculty of 1000 Medicine”(千名医学家)数据库。其中,“千名生物学家”根据 1400 多位专家学者的建议,提供目前世界上最重要的生物学论文信息

及研究趋势;“千名医学家”则来自美国哈佛大学和英国剑桥大学等著名学府的约 2500 名顶级医学专家将可能改变医学实践的极少数优秀论文推荐给大家。上述系统以事后评议的方式(post publication peer review,简称 PPR),采用加权平均的方法将专家意见汇总,计算出每篇论文的 F1000 因子(F1000 Article Factor,下文简称 FFa),其中 6 分为推荐论文,8 分为必读论文,10 分为杰出论文。作为本文比较对象的 Web of Science,则是世界权威的科学技术文献索引工具,其独特的引文检索体系,使人们不仅可以迅速地组建研究课题的参考文献网络,更为重要的是,可以从文献引证的视角评估文章的学术价值。以上述两个数据库为代表,从相关性视角进行同行评议和文献计量关系研究应该具有典型意义。

2 数据的采集与统计

文中将回溯年限确定为 2005 年。为此,通过 F1000 高级检索界面,笔者选取了该数据库中免疫学、生物信息学为统计对象。检索式为: article publication data = 2005 and subject = “immunology” or subject = “bioinformatics”。检索结果为免疫学文献 981 篇,生物信息学文献 999 篇,共计 1980 篇文献。对于上述记录,首先采用计算机编程抓取网页中的篇名、刊名、作者、FFa 等信息,并将这些信息绘制成 Excel 表格;继而利用 Web of Science 数据库,通过篇名检索每篇文章的被引量,对于相同主题文献,通过作者、发表年限、刊名等进一步加以区分与确认,以避免误检,统计区间为 2005 年至 2011 年 4 月,检索及统计如下^[10]:

981 篇免疫学文献中,978 篇(占总论文量的 99.7%)为 Web of Science 收录;FFa 最高值为 26 分,该文刊登在《细胞》杂志上;10 分以上的优秀论文 198 篇,占 20.2%;8 分以上的必读论文为 247 篇。上述文献来自 136 种期刊,按照载文量排序,依次为《实验医学学报》(*Journal of Experimental Medicine*)、《免疫学期刊》(*Journal of Immunology*)、《美国科学院院报》(*Proceedings*

of the National Academy of Sciences of the United States of America)、《自然免疫学》(*Nature Immunology*)、《科学》(*Science*)。

999 篇生物信息学论文中,989 篇(占总论文量的 99%)为 Web of Science 所收录;FFa 最高值为 62 分,该文刊登在《自然》杂志上;10 分以上的优秀论文为 173 篇,占 17.3%;8 分以上的必读论文为 259 篇。上述论文来自于 181 种期刊,依据载文量依次为《美国科学院院报》、《科学》、《自然》等。

3 数据的处理

依据上述统计结果,借助 SPSS16.0 将分别对两个学科的 FFa 与被引频次的相关性进行统计学检验。因分析中需剔除存在缺失值的记录,鉴于 WoS 中的收录情况,将有 978 篇免疫学文献、989 篇生物信息学文献参与该部分分析。

3.1 数据的正态分布检验

为了选择适宜的统计检验方法,以生物信息学为例,首先采用 K-S 检验对数据进行正态性分析(见表 1),分析结果表明显著性水平小于 0.05,因此无论是 FFa 还是 WoS 中被引频次均不符合正态分布,应选取非参数的斯皮尔曼检验为宜^[11]。

表 1 生物信息学数据的正态性检验结果

项目	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FFa	.286	999	.000
被引频次	.274	989	.000

3.2 FFa 与被引频次相关性分析

989 篇生物信息学文献的 FFa 与 WoS 中被引频次的斯皮尔曼检验结果如表 2 所示。

978 篇免疫学文献的 FFa 与 WoS 中被引频次的斯皮尔曼检验结果如表 3 所示。

表 2 生物信息学文献的 FFa 与 WoS 中被引频次相关性检验结果

			FFa	被引频次
Spearman's rho	FFa	Correlation Coefficient	1.000	.415 **
		Sig. (2-tailed)		.000
		N	999	989
	WoS 被引频次	Correlation Coefficient	.415 **	1.000
		Sig. (2-tailed)	.000	
		N	989	989

** Correlation is significant at the 0.01 level(2-tailed).

表 3 免疫学文献 FFa 与 WoS 中被引频次相关性检验结果

			FFa	被引频次
Spearman's rho	FFa	Correlation Coefficient	1.000	.440 **
		Sig. (2-tailed)		.000
		N	981	978
	WoS 被引频次	Correlation Coefficient	.440 **	1.000
		Sig. (2-tailed)	.000	
		N	978	978

** Correlation is significant at the 0.01 level(2-tailed).

由表 2、表 3 看出,在 0.01 的显著性水平上,生物信息学文献 FFa 与 WoS 中被引频次的相关系数为 0.415,免疫学文献 FFa 与 WoS 中被引频次的相关系数为 0.440。因此,无论是生物信息学还是免疫学,上述两个变量都显著相关,且为正相关。

4 结果与讨论

根据统计学知识,如果用 r 代表相关系数,当 $0 < r \leq 1$ 时,表明变量之间存在正相关关系,即两个变量变动方向相同,同时依据变量之间的相关程度,将相关程度分为高度相关($|r| \geq 0.8$)、中度相关($0.5 \leq |r| < 0.8$)、低度相关($0.3 \leq |r| < 0.5$)、不相关($|r| < 0.3$)四种情形^[12]。依据上述标准,无论是生物信息学还是免疫学文献,其 FFa 与被引频次虽然正相关,但均属低度相关,由此证明两个变量在变动方向一致的主流趋向下,也存在着相异之处。

为了进一步分析导致这种差异的原因,本文从两个方面对原始数据进行挖掘。首先析出 FFa 与被引频次关系中差异最大的 50 篇文献,进而按照同样的方法找出差异最小的 50 篇文献。

4.1 FFa 与被引频次差异最大的 50 篇文献
这里以生物信息学为例,初步分析表明,当按照 FFa、被引频次分别排序时,各自排序中的前 50 篇文献仅有 12 篇相同,值得注意的是,FFa 与被引频次最大值不在 12 篇范畴之中。居于 FFa 首位的文章(FFa 为 62)对应被引频次为 87 次,位于被引的第 325 位;与此同时被引频次排序表明,在统计期间内最高值为 2342 次,该文 FFa 为 10,FFa 排序为 110 位。那么,我们能否通过统计分析找出差异最大的 50 组数据呢?

由于 FFa 与被引频次的不可比性,所以标准化是必要的前期准备工作。为此本文采用 Z 分数与 T 分数将两组数据转化成方差为 10、平均分 50 的标准值,进而将两组数据相减,并将相减后结果的绝对值进行比较,从而筛选出前 50 组。检验表明,在 0.01 的显著性水平上,50 组数据的相关系数为 -0.416,因此在上述 50 篇文献中,FFa 与被引频次存在负向相关关系,这就意味着存在专家打分很高而被引较低,或者被引较高而专家打分很低的情形。为了清晰地体现这种负向关系,表 4 罗列了该 50 篇文献的具体信息。

表 4 FFa 与被引频次差异最大的 50 篇文献

序号	篇 名	刊 名	FFa	被引频次
1	Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis	Nature	62	87
2	A haplotype map of the human genome	Nature	10	2342
3	Gene set enrichment analysis:a knowledge-based approach for interpreting genome-wide expression profiles	Proc Natl Acad Sci U S A	8	1755
4	Genome sequencing in microfabricated high-density picolitre reactors	Nature	18	1704
5	Combinatorial microRNA target predictions	Nat Genet	7	1015
6	Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles	Science	27	278
7	The map-based sequence of the rice genome	Nature	7	962
8	Pathogenic fungus harbours endosymbiotic bacteria for toxin production	Nature	23	96
9	MicroRNA expression profiles classify human cancers	Nature	19	1635
10	Complement factor H polymorphism and age-related macular degeneration	Science	6	794

续表

序号	篇 名	刊 名	FFa	被引频次
11	Complement factor H variant increases the risk of age-related macular de- generation	Science	6	771
12	Cytokinin oxidase regulates rice grain production	Science	23	224
13	Silencing of microRNAs in vivo with “antagomirs”	Nature	10	842
14	Recurrent fusion of TMPRSS2 and ETS transcription factor genes in pros- tate cancer	Science	8	723
15	An aneuploid mouse strain carrying human chromosome 21 with Down syn- drome phenotypes	Science	18	107
16	Interchromosomal associations between alternatively expressed loci	Nature	21	285
17	A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia	N Engl J Med	8	691
18	Complement factor H polymorphism in age-related macular degeneration	Science	17	1204
19	The transcriptional landscape of the mammalian genome	Science	9	729
20	Logic of the yeast metabolic cycle;temporal compartmentalization of cellular processes	Science	18	165
21	Towards a proteome-scale map of the human protein-protein interaction net- work	Nature	10	770
22	Algae acquire vitamin B12 through a symbiotic relationship with bacteria	Nature	16	69
23	Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome	Proc Natl Acad Sci U S A	16	84
24	A universal trend of amino acid gain and loss in protein evolution	Nature	16	84
25	Systematic discovery of regulatory motifs in human promoters and 3’ UTRs by comparison of several mammals	Nature	11	788
26	Phylogenetic shadowing and computational identification of human microR- NA genes	Cell	6	484
27	Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system	Nature	17	173
28	Genome-wide map of nucleosome acetylation and methylation in yeast	Cell	6	469
29	Plectasin is a peptide antibiotic with therapeutic potential from a saprophyt- ic fungus	Nature	16	126
30	Assessing computational tools for the discovery of transcription factor bind- ing sites	Nat Biotechnol	6	462
31	Aging and death in an organism that reproduces by morphologically sym- metric division	PLoS Biol	16	129

续表

序号	篇 名	刊 名	FFa	被引频次
32	An anaerobic mitochondrion that produces hydrogen	Nature	15	74
33	Self-organized patchiness in asthma as a prelude to catastrophic shifts	Nature	15	88
34	A gene expression map of Arabidopsis thaliana development	Nat Genet	10	675
35	RNA-dependent cysteine biosynthesis in archaea	Science	15	94
36	No transcription-translation feedback in circadian rhythm of KaiC phosphorylation	Science	16	152
37	The selective cause of an ancient adaptation	Science	14	37
38	Epigenetic differences arise during the lifetime of monozygotic twins	Proc Natl Acad Sci U S A	9	607
39	Functional genomic analysis of the Wnt-wingless signaling pathway	Science	16	163
40	Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes	Genome Res	10	654
41	Nova regulates brain-specific splicing to shape the synapse	Nat Genet	16	170
42	Gene transfer to plants by diverse species of bacteria	Nature	14	56
43	The RETINOBLASTOMA-RELATED gene regulates stem cell maintenance in Arabidopsis roots	Cell	15	114
44	Reproducible research;a bioinformatics case study	Stat Appl Genet Mol Biol	13	9
45	Structures of complement component C3 provide insights into the function and evolution of immunity	Nature	15	125
46	Initial sequence of the chimpanzee genome and comparison with the human genome	Nature	11	696
47	TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing	Nature	6	405
48	A high-resolution map of active promoters in the human genome	Nature	6	400
49	Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast	Nat Cell Biol	14	81
50	An index to quantify an individual's scientific research output	Proc Natl Acad Sci U S A	12	732

4.2 FFa 与被引频次差异最小的 50 篇文献

出于比较与论述的需要,我们以同样的方法获得了差异最小的 50 组数据。统计检验表明,在 0.01 的显著性水平上,50 组数据的相关

系数为 0.702,在此,FFa 与被引频次呈现出正相关性。数据深入分析显示,这 50 组数据 FFa 值均列入 10 分以下,被排除在优秀论文之外,其中 42 篇(占 84%)属于 8 分以下的推荐论文,而 40

篇(占80%)年平均被引量在1.5次以下,与最高值371次相去甚远。50组数据具体信息从略。

4.3 结果分析

FFa与被引频次的正向相关性以及50组差异最小的数据的析出都说明一个事实:一些论文既被同行高度认可,也高频被引;另外一些论文专家只作为一般推荐,其被引也很低,即Félix de Moya所谓定量测量与专家意见始终保持一致的理想状态^[13-14]。与此同时50组差异最大的数据则表征着引文分析与同行评议意见相左情形的存在:有一些被专家打分很高的论文并没有高频被引,反之,一些高频被引的论文专家打分并不高。

究其原因,虽然FFa与被引频次是同一问题的两个方面,但是差异是客观存在的。从根本上讲,FFa作为同行评议的代表反映的是质量,而被引作为文献计量的经典指标借助数量反映的是影响;就机理而言,FFa是固定不变的,而被引,特别是高频被引的文献,由于受到马太效应的影响,随着时间的推移再次被引的几率会增加,因而被引具有随年代变化的特性;从时间来说,同行评议与论文的发表几乎是同步进行的,而被引无疑是滞后的,并且受到半衰期等因素的影响,引文峰值出现时间因此会各不相同;从内容来看,50篇优秀论文入选缘由统计表明,高的FFa值是与里程碑性的文章联系在一起的,它们几乎都是New Finding(新奇的发现)、Technical Advance(技术的进展)、Interesting Hypothesis(有趣的假说)、Important Confirmation(重要的确认)和Controversial Findings(争议性的发现)的结合体,而高频被引的文献以综述性文章居多。上述原因造成了同行评议与文献计量总体一致情况下的部分相异。

因此,无论是同行评议亦或文献计量,单独作为科研评价的方法都会有失偏颇。进一步而言,在现今科学评价以非此即彼的二元制占据主导的情形下,不能武断地作出同行评议与文献计量孰优孰劣的判断。英国高等教育资助委员会以文献计量与同行评议正向相关作为改革

的起点,最终一改其完全实施文献计量的初衷,而达成以同行评议为主、引文分析为辅的实施细则也许是最有力的例证。因此我们所能指出的是,无论是F1000还是WoS都有待完善,那么目前将二者有机结合将是较为全面审视科学评价的有力措施。

5 结论

通过上述统计与分析,得出结论:

(1)总体而言,一定程度上F1000与WoS提供了一致的结论。从统计学的视角来看,文献计量指标与同行评议结果具有正向相关性,这为文献计量在科学评价中的有效性提供了借鉴,同时也为科研评价从定性分析到定量评估的转化提供了有力的依据。

(2)F1000、WoS都有局限性,因此现今文献计量指标仍不足以单独作为评价标准,如果仅仅依赖于文献计量指标,会遗漏某些刊载重要成果的论文,而这些或许恰恰是专家确认的优秀论文。同样就分析结果来看,同行评议单独作为科学评价的方法,仍然有其不完善之处,因此将同行评议作为判断标准沿用的同时,应辅以文献计量指标。

(3)科学评价须脱离传统的非此即彼的二元制,文献计量与同行评议是相辅相成的,并非取而代之的关系。以引文分析为代表的定量指标与同行评议方法的结合将是未来科学评价的主流。

参考文献:

- [1] Oppenheim C. Out with the old and in with the new: The RAE, bibliometrics and new REF[J]. Journal of Librarianship and Information Science, 2008, 40(3): 147-149.
- [2] Van Noorden R. A profusion of measures[J/OL]. Nature, 2010, 465: 864-866 [2011-05-13]. <http://www.nature.com/news/2010/100616/full/465864a.html>.
- [3] Norris M, Oppenheim C. Citation counts and the research assessment exercise V: Archaeology and

- the 2001 research assessment exercise[J]. *Journal of Documentation*, 2003, 59(6): 709 – 730.
- [4] So C. Citation ranking versus expert judgment in evaluating communication scholars: Effects of research specialty size and individual prominence[J]. *Scientometrics*, 1998, 41(3): 325 – 333.
- [5] Norris M, Oppenheim C. Peer review and the h-index: Two studies[J]. *Journal of informetrics*, 2010, 4(3): 221 – 232.
- [6] Patterson M S, Harris S. The relationship between reviewers' quality – scores and number of citations for papers published in the journal physics in medicine and biology from 2003 – 2005[J]. *Scientometrics*, 2009, 80(2): 343 – 349.
- [7] Derrick G E. The association between four citation metrics and peer rankings of research influence of Australian researchers in six fields of public health[J/OL]. *Plos One*, 2011, 6(4): e18521 [2011 – 08 – 18]. <http://www.ncbi.nlm.nih.gov/pubmed/21494691>.
- [8] David A W. Do 'Faculty of 1000' (F1000) ratings of ecological publications serve as reasonable predictors of their future[J]. *Ideas in Ecology and Evolution*, 2010, 3: 11 – 15.
- [9] F1000[OL]. [2011 – 03 – 30]. <http://f1000.com/>.
- [10] Web of knowledge[OL]. [2011 – 04 – 30]. <http://isiknowledge.com/>.
- [11] [英] 迈克·塞沃尔. 链接分析: 信息科学的研究方法[M]. 孙建军, 译. 南京: 东南大学出版社, 2009: 34. (Mike Thelwall: *Link Analysis: An Information Science Approach* [M]. Sun Jianjun, translate. Nanjing: Southeast University Press, 2009: 34.)
- [12] 薛薇. SPSS 统计分析方法及应用[M]. 北京: 电子工业出版社, 2009: 245. (Xue Wei. *Statistical Analysis Methods and Applications for SPSS* [M]. Beijing: Electronic Industry Press, 2009: 245.)
- [13] Greene S. Peer review and the age of aquarius: It's time to reinvent the system that validates scientific discovery[J/OL]. *The Scientist*, 2010, 24(8): 13 [2011 – 05 – 15]. <http://f1000scientist.com/article/display/57580/>.
- [14] Primož Južni č. Scientometric indicators; Peer-review, bibliometric methods and conflict of interests[J]. *Scientometrics*, 2010, 85(2): 429 – 441.

宋丽萍 天津师范大学管理学院副教授。通讯地址: 天津市天津师范大学管理学院。邮编: 300387。

王建芳 中国科学院国家科学图书馆副研究员。通讯地址: 北京市中关村北四环西路 33 号。邮编: 100190。

(收稿日期: 2011-07-01; 修回日期: 2011-09-02)