

H 指数在期刊同行评议质量评价中的应用*

石俊强 林琳 甘慧敏

(《济宁医学院学报》编辑部, 济宁 272013)

摘要 期刊编辑通常依赖于同行审稿专家的建议,对作者提交的学术研究、想法或材料做出决策。同行审稿专家为编辑提供稿件质量的评估,为作者提供建设性的意见。超过 80% 的受访学者认为期刊同行评议有助于科学交流以及学术期刊质量的提升。然而,同行评议本身存在许多缺点,目前对于如何评价期刊同行评议的质量和有效性没有统一的评价标准。本文重点介绍 H 指数在同行评议质量评价中的应用,为我国期刊同行评议质量评价提供有价值的新视角。

关键词 同行评议;质量评价;H 指数

中图分类号:G237.5 文献标识码:A 文章编号:1000-9760(2019)12-441-04

Application of H-index in the evaluation of peer review

SHI Junqiang, LIN Lin, GAN Huimin

(Editorial Department of Journal of Jining Medical University, Jining 272013, China)

Abstract: Journal editors usually rely on the views of independent experts (peers) in making decisions on an author's scholarly work, research, or ideas and material submitted for publication. The peer reviewer serves the editor by substantiating the quality of the manuscript, and serves the author by giving constructive criticism. More than 80% of surveyed academics agreed that journal peer review greatly helps scientific communication and improves the quality of academic journals. However, there are a lot of shortcomings in peer review, and there appears to be little agreement about how to measure its quality and effectiveness. This paper focuses on the application of H-index in the evaluation of peer review, which will provide a new perspective for the evaluation of the quality of peer review.

Keywords: Peer review; Evaluation of the quality; H-index

同行评议起源于 17 世纪的欧洲,20 世纪末已成为国际学术期刊出版工作中的一个重要质量控制机制,其中《Nature》杂志 1967 年开始采用同行评议^[1]。英国社会科学院发布的一份报告显示同行评议制度是控制科学研究质量的最佳方式^[2]。2009 年, Publishing Research Consortium 对全球 3000 多个学术机构做了一项关于同行评议的调查,85% 学术机构认为同行评议极大地促进了科学交流,83% 学术机构认为没有同行评议科学研究质量将无法保障,90% 学术机构认为同行评议有效地提高了发表论文的质量^[3]。目前期刊编辑通常依赖于同行审稿专家的建议,对作者提交的学术研

究、想法或材料做出决策。同行审稿专家为编辑提供稿件质量的评估,为作者提供建设性的意见。由此可见,同行评议是学术期刊出版过程中特别重要的一环,它能提高发表论文的质量、可读性。

目前,各种学术期刊为保障同行评议的质量,普遍使用单盲评审、双盲评审、开放式评审等形式。然而,同行评议依然存在许多缺陷,如同行评议制度中可能存在因经济利益、社会关系、竞争关系、个人好恶而导致的非公正性,因评价标准和程序的机械化、专业歧视、热点追逐导致的非合理性,因虚假陈述、能力缺陷、怠于勤勉而导致的非客观性^[4-6]。开放存取(Open Access, OA)自出现以来,OA 期刊得以迅速发展,目前全球已有 9000 多种 OA 期刊,其中大部分像《PLoS One》一样采用基于作者付费(author publication charge, APC)的开放存取模式,

* [基金项目] 济宁医学院青年教师科研扶持基金(JY2017 RW012)

而 APC 模式的应用降低了出版商的编辑标准,同行评议质量也大大降低^[7]。2009 年,《The Open Information Science Journal》杂志发表了一篇文章,而该篇文章作者把数据、图表、相关参数及参考文献输入自己建立的一个软件程序中,生成了这篇看似合理却没有任何科学意义的论文^[7]。同行评议在这次审稿中没有保障发表论文的真实性、科学性。因此,基于 APC 模式 OA 期刊采用的同行评议同样存在许多缺陷,许多专家呼吁基于客观的标准,如在文献计量指标对同行评议的质量进行评估。目前对于如何评价期刊同行评议的质量和有效性没有统一的评价标准。本文将重点介绍 H 指数在同行评议质量评价中的应用,为我国期刊同行评议质量评价提供有价值的新视角。

1 H 指数

2005 年,美国物理学家 Hirsch 提出了 H 指数,用于评价研究人员的学术成就^[8]。目前,H 指数已成为信息计量学和学术评价研究的前沿领域之一。某研究人员的 H 指数是指他或她有 h 篇论文分别被引用了至少 h 次。即在某数据库中让论文按被引次数依次从高到低排列,直到某篇论文的序号大于该论文被引次数,那个序号减去 1 就是 H 指数。Braun 等进一步拓宽了 H 指数的应用,他们将 H 指数用于期刊学术影响力评价中,并作出新的定义:对于一种期刊,如果它发表的全部论文中有 H 篇文章,每篇被引用数至少为 H,同时要满足 H 这个自然数为最大,那么 H 即为该期刊的 H 指数^[9]。

有一些研究表明,较早被聘为审稿人的审稿专家比较晚聘为审稿人的审稿专家可能获得更高的 H 指数,给出更高质量的同行评议^[7,10-11]。因此,H 指数也可作为一个指标用于同行评议质量的评价。

2 H 指数在同行评议质量评价中的应用

2.1 同行评议过程中的参与者

2.1.1 主编 主编是某一出版物的领航者,在编委会或总编辑领导下,按编辑方针实施某个方面的具体业务。其工作内容为研究、制订某一方面的工作计划,组织并领导日常编辑事务,如审稿、审样、撰写重要言论等。有些编辑部门不设总编辑,主编是该部门的负责人,其职权相当于总编辑。主编通常按同行评议的结果决定稿件是否适合在该刊发表。

2.1.2 副主编 通常期刊社所收到稿件的全面审查和筛选由副主编负责。如果稿件适合期刊社的发表并需要进一步审查,副主编会选择合适的审稿专家并将稿件发给审稿专家进行同行评议。同时副主编还指导和帮助编辑人员解决工作中的疑难问题,参与组稿、审稿、校对工作,必要时对某些重要稿件进行审理和修改,对争议稿件约请主编和编委会编委复议,并承担审稿稿件的保密责任。

2.1.3 审稿专家 如果把主编和副主编比作一艘船舶的驾驶员和副驾驶,那么审稿专家就是这艘船舶的引擎。无论传统的学术出版方式还是开放获取的出版方式,学术期刊都需要评审专家对已提交的文章进行评审和评价。大多数期刊把稿件发送给 2~4 名同行审稿专家。审稿专家从专业角度鉴别稿件质量,衡量稿件的科学性、真实性、创新性、可行性等;判断稿件的可发表性,为编辑提供客观、科学的决策依据;提出建设性的修改意见,帮助作者完善研究内容,提高论文质量;防范学术造假,在一定程度上评审专家能发现抄袭或伪造的研究^[12]。

2.2 同行评议评分(Peer Review Evaluation Score, pre-SCORE)^[7]

pre-SCORE 基本公式 $S = [A + B + C] / \sqrt{V}$

pre-SCORE 延伸公式 $S = [X * E + Y * F + Z * G] / \sqrt{V}$ 其中 S 表示 pre-SCORE, $A = X * E$, $B = Y * F$, $C = Z * G$, X 表示处理某篇稿件主编参与人数, Y 表示处理该篇稿件副主编参与人数, Z 表示处理该篇稿件审稿专家参与人数, E (一般等于 0.4) 表示主编处理该篇稿件所做贡献的权重, F (一般等于 0.3) 表示副主编处理该篇稿件所做贡献的权重, G (一般等于 0.2) 表示审稿专家处理该篇稿件所做贡献的权重, V 表示正在审阅稿件的版本数(原稿计为 0, 修订版 1 计为 1, 修订版 2 计为 2, 依次类推)。

2.3 H 指数往 pre-SCORE 公式的导入及应用

2.3.1 含有 H 指数的 pre-SCORE 扩展公式 用 $R_{e1,2,3,\dots,n}$ 代表每个主编的 H 指数, $R_{a1,2,3,\dots,n}$ 代表每个副主编的 H 指数, $R_{r1,2,3,\dots,n}$ 代表每个审稿专家的 H 指数,把参与稿件审查人员的 H 指数带入 pre-SCORE 公式中,则 pre-SCORE 扩展公式 $S = [(R_{e1} * A) + (R_{a1} * B) + (R_{r1} * C) + (R_{e2} * C) + \dots + R_m * C] / \sqrt{V}$

2.3.2 举例说明 目前大多数学术期刊编辑部都

采用在线投稿与审稿系统,在线投稿与审稿系统中包含主编、副主编、责任编辑、审稿专家等的信息,这为同行评议质量评价提供了便利。现以一篇稿件为例,这篇稿件 2011 年 1 月投到在线投稿与审稿系统中,12 月采用发表,中间稿件送往 3 位审稿专家进行同行评议,稿件经历 2 次修改 3 轮审查^[7]。这篇稿件主编的 H 指数(H 指数的测定可利用汤姆森路透社知网)为 36,副主编的 H 指数为 52,3 位审稿专家的 H 指数分别为 45、39 和 28,则稿件第一轮审查 pre-SCORE 为

$$\begin{aligned} S_1 &= [(R_{e1} * A) + (R_{a1} * B) + (R_{r1} * C) \\ &\quad + (R_{r2} * C) + (R_{r3} * C)] / \sqrt{V} \\ &= [(36 * 0.4) + (52 * 0.3) + (45 * 0.2) \\ &\quad + (39 * 0.2) + (28 * 0.2)] / \sqrt{1} \\ &= [14.4 + 15.6 + 8.4 + 9.0 + 5.6] / 1 \\ &= 53.0 \end{aligned}$$

本稿件被送回作者修改后重新提交,所有同行评议过程中的参与者再次评估文章,此时 \sqrt{V} 变为 $\sqrt{2}$,其他变量保持不变,则稿件第二轮审查 pre-SCORE 为

$$\begin{aligned} S_2 &= [(36 * 0.4) + (52 * 0.3) + (45 * 0.2) \\ &\quad + (39 * 0.2) + (28 * 0.2)] / \sqrt{2} \\ &= [14.4 + 15.6 + 8.4 + 9.0 + 5.6] / 1.4 \\ &= 37.9 \end{aligned}$$

本稿件将返回到作者再次修改和提交。副主编再次审核文章,对所有的同行评议和作者修改感到满意后,再提审主编。主编回顾以前所有的同行评议,重新阅读文章和决定接受文章。则稿件第 3 轮审查 pre-SCORE 为

$$\begin{aligned} S_3 &= [(36 * 0.4) + (52 * 0.3)] / \sqrt{3} \\ &= [14.4 + 15.6] / 1.7 \\ &= 17.6 \end{aligned}$$

稿件每一轮审查都需要计算 pre-SCORE。这个例子中稿件 pre-SCORE 的总分为

$$\begin{aligned} S &= S_1 + S_2 + S_3 \\ &= 53.0 + 37.9 + 17.6 \\ &= 108.5 \end{aligned}$$

2.4 应用

2.4.1 实际应用的可行性 目前大多数学术期刊编辑部都采用在线投稿与审稿系统,通过在线投稿与审稿系统可以捕获所有参与同行评议过程的参与人员的信息。此信息被标记在系统 XML 中,并且可以通过创建的软件来导出和处理。

2.4.2 学术期刊单篇稿件 pre-SCORE 的计算 利用汤姆森路透社知网、知网、万方等查询或计算某学术期刊主编、副主编、审稿专家的 H 指数,带入 pre-SCORE 公式,计算该篇稿件的 pre-SCORE。

2.4.3 学术期刊每一期学术水平的预测 除了为每一篇论文提供一个 pre-SCORE,每期发行的期刊学术水平可通过计算该期所有论文 pre-SCORE 总分的平均值来预测。

2.4.4 学术期刊每一卷(或年)学术水平的预测 根据 3.4.3 计算方法,同理也可以用于学术期刊每一卷(或年)学术水平的预测。

3 讨论

同行评议是当前学术期刊普遍采用的评估稿件质量、遴选高质量稿件的手段,在控制期刊的学术质量、规范学者的学术行为方面发挥了重要作用^[2-3]。然而,同行评议本身存在许多缺陷,许多专家呼吁基于客观的标准,如在文献计量指标对同行评议的质量进行评估。目前对于如何评价期刊同行评议的质量和有效性没有统一的评价标准。

目前,H 指数已成为信息计量学和学术评价研究的前沿领域之一。H 指数首先应用于研究人员学术水平的评价。Braun 等^[9]拓宽了 H 指数的应用,他们将 H 指数用于期刊学术影响力评价中,而 Etkin 等^[7]也将 H 指数用于期刊单篇稿件的 pre-SCORE 的计算中。目前大多数学术期刊编辑部采用在线投稿与审稿系统,通过汤姆森路透社网、知网、万方等可以捕获所有参与同行评议过程的参与人员的信息(如期刊主编、副主编、审稿专家的 H 指数)。此信息被标记在系统 XML 中,并且可以通过创建的软件来导出和处理,这为计算期刊单篇稿件的 pre-SCORE 提供了便利。H 指数能够客观评价研究人员的学术成就,同样也能评价期刊主编、副主编、审稿专家的学术水平,因此,H 指数应用于稿件的 pre-SCORE 的计算中能够客观反映同行评议质量水平。有一些研究表明,较早被聘为审稿人的审稿专家比较晚聘为审稿人的审稿专家可能获得更高的 H 指数,给出更高质量的同行评议^[7,10-11]。

虽然 H 指数作为一个指标可用于同行评议质量的评价,但不能全面反映同行评议的质量。从 pre-SCORE 公式可以看出,pre-SCORE 对同行评议的内容本身没作评估。Ausloos 等^[13]应用 Zipf 定

律对同行评议的内容本身进行评估,Zipf 定律是描述词频分布规律的强大数学工具;作为经验定律,它对同行评议的内容本身进行评估仍有不足之处,有待进一步完善。如果把 H 指数和 Zipf 定律合理组合应用于同行评议质量的评价,H 指数和 Zipf 定律将扬长避短,取长补短,发挥各自的优势,能客观全面的评价同行评议的质量,pre-SCORE 公式可描述为 $S = [(R_{e1} * A) + (R_{a1} * B) + (R_{r1} * C) + (R_{r2} * C) + \dots + R_m * C] / \sqrt{V} + Z$ 。然而,H 指数和 Zipf 定律应用于同行评议质量的评价过程中,各自在 pre-SCORE 公式占得权重是多少需要我们进一步去研究,另外,应用 H 指数和 Zipf 定律的 pre-SCORE 分数应以百分制为宜,这样同行评议质量的评价可以根据分数分为优、良、差等分级。

参考文献:

[1] Leopold SS. Editorial: peer review and the editorial process--a look behind the curtain[J]. Clin Orthop Relat Res, 2015, 473 (1): 1-3. DOI: 10. 1007/s11999-014-4031-x.

[2] Radnofsky L. Learned society defends peer review[R]. Times Higher Education, 2007.

[3] Publishing Research Consortium. Peer review in scholarly journals: perspective of the scholarly community—an international study[R]. Mike Monkman Media, 2009.

[4] Black N, Van Rooyen S, Godlee F, et al. What makes a good reviewer and a good review for a general medical

journal[J]. JAMA J Am Med Assoc, 1998, 280(3): 231 - 233.

[5] Bohannon J. Who's afraid of peer review[J]. Sci Mag. 2013, 342(6154): 60-65.

[6] 常唯, 曹会聪, 曹金, 等. 国际科技期刊同行评议的实践与特点[J]. 中国科技期刊研究, 2016, 27(1): 10-17. DOI: 10. 11946/cjstp. 201509050854.

[7] Etkin A. A new method and metric to evaluate the peer review process of scholarly journals[J]. Pub Res Q, 2014, 30: 23 - 38. DOI 10. 1007/s12109-013-9339-y.

[8] Hirsch JE. An index to quantify an individual's scientific research output[J]. Proc Nat Acad Sci, 2005, 102(46): 16569-16572.

[9] 姜晓. 构建西部图书馆信息服务网络模式的设想[J]. 图书情报工作, 2005, 49(3): 76-79.

[10] Black N1, van Rooyen S, Godlee F, et al. What makes a good reviewer and a good review for a general medical journal[J]. JAMA, 1998, 280(3): 231-233.

[11] Callaham M1, McCulloch C. Longitudinal trends in the performance of scientific peer reviewers[J]. Ann Emerg Med, 2011, 57(2): 141-148. DOI: 10. 1016/j. annemergmed. 2010. 07. 027.

[12] 周沫, 王维朗. 科技期刊同行评审专家与编辑的博弈[J]. 编辑学报, 2013, 25(3): 214-216.

[13] Ausloos M, Nedic O, Fronczak A, et al. Quantifying the quality of peer reviewers through Zipf's law[J]. Scientometrics, 2016, 106: 347-368. DOI: 10. 1007/s11192-015-1704-5.

(收稿日期 2019-11-09)

(本文编辑:石俊强)

• 读者 • 作者 • 编者 •

本刊对来稿中统计学处理的有关要求

1. 统计研究设计: 应交待统计要就设计的名称和主要做法, 如调查设计具体类型; 实验设计类型等。主要围绕着 4 个基本原则(随机、对照、重复、均衡)概要说明, 尤其是要交待如何控制非试验因素的干扰和影响。

2. 资料的表达与描述: 用 $\bar{x} \pm s$ 表达近似服从正态分布的定量资料, 用 M(QR) 表达呈偏态分布的定量资料; 用统计表时, 须合理安排纵横坐标, 并将数据的意义表达清楚; 用统计图时, 所用统计图的类型应与资料性质匹配, 并使数轴上刻度值的标法符合数字原则; 用相对数时, 分母不宜小于 20, 须注意区分百分率与百分比。

3. 统计学分析方法: 对于定量或定性资料, 应根据所采用的设计类型、资料所具备的条件和分析目的, 选择合适的统计学分析方法, 不应盲目套用 t 检验、单因素方差分析或 χ^2 检验; 对于回归分析, 应结合专业知识和散布图, 选用合适的回归类型; 对具有重读实验数据的回归分析资料, 不应简单化处理, 对于多因素、多指标资料, 要在一元分析的基础上, 尽可能运用多元统计学分析方法, 以便对因素之间的交互作用和对指标之间的内在联系进行全面、合理的解释和评价。

4. 统计结果的表达: 当 $P < 0.05$ (或 $P < 0.01$) 时, 应说明对比组之间的差异有统计学意义, 而不应说对比组之间具有显著性(或非常显著性)的差别; 应写明所用统计学分析方法的具体名称(如: 成组设计资料的 t 检验、两因素析因设计资料的方差分析、多个均数之间两两比较的 q 检验等), 统计量的具体值(如 t 值, χ^2 值, F 值等), 应尽可能给出具体的 P 值; 当涉及总体参数(如总体均数、总体率等)时, 在给出显著性检验结果的同时, 再给出 95% 可信区间。

(本刊编辑部)