试析应用SCI数据进行科研评估的局限性与建议

王晓丽 孙洁 刘洁
(中国科学技术信息研究所, 北京 100038)

摘要: 针对目前我国科研评估中过度倚重SCI数据的问题, 结合实际工作, 从SCI学科语种的差异性、地址数据缺陷、作者顺序问题以及被引频次和影响因子等方面对其在科研评估中的局限性进行分析。研究认为, 只有通过制定有导向性的政策和多角度的评估方案, 提高对科研产出的管理意识, 完善SCI数据存在的缺陷, 规范评估检索流程等措施, 才能为科研评估中所涉及到的问题提供公平、公正、规范的解决方案。

关键词: SCI; 影响因子; 科研评估; 被引频次; 学科差异

中图分类号: G250.7

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2018.05.016

Analysis of Limitations in Scientific Research Evaluation Based on SCI Data

WANG Xiaoli, SUN Jie, LIU Jie

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: In view of the problem of SCI data in China's scientific research evaluation, this paper combines with the actual work, from the differences in SCI language, address data defects, author order issues, the citations and impact factors to analyze its limitations in scientific evaluation. In order to solve the problems and provide fair, just and standardized solutions, it is necessary to establish a guiding policy and a multi-angle evaluation plan; improve the management awareness of scientific research, correct the defects of SCI data, standardize retrieval process, etc.

Keywords: science citation index, impact factor, scientific research evaluation, citation frequency, disciplinary difference

SCI (Science Citation Index) 数据库是由美国科学信息研究所于1961年创办出版的科学引文数据库, 它收录了150个自然科学学科的8000多种主要期刊, 同时还收录了论文所有引用的参考文献, 是目前国际公认的科技论文主要检索工具, 在进行科学统计与科学评价中具有一定的全面性和权威性^[1]。

由于SCI收录世界各国自然科学领域所有最新研究成果, SCI论文越来越受到国内外科研人

员、政府的科技决策部门和科学基金资助部门的重视和关注。国际上许多国家的科学计量机构在科研绩效评估中, 常用SCI数据库作为统计源以定量研究科研能力、科学发展趋势与活动水平^[2]。

1 SCI在我国和主要国家科研评估中的应用

在我国, SCI应用于评估科学研究水平始于上世纪80年代。随着我国科学技术水平的飞速发展, 以及国家对科研投入的持续增加, 我国被

作者简介: 王晓丽 (1981—), 女, 图书馆学硕士, 中国科学技术信息研究所助理研究员, 研究方向: 情报分析方法 (通讯作者); 孙洁 (1980—), 女, 硕士, 中国科学技术信息研究所助理研究员, 研究方向: 科技查新; 刘洁 (1980—), 女, 硕士, 中国科学技术信息研究所助理研究员, 研究方向: 科技查新。

收稿时间: 2018年6月13日。

SCI收录的科研论文数量逐年增加。同时由于SCI的引导作用，很多科研工作者积极向高水平期刊投稿，让自己的科研成果接受更高水平的同行专家评议，从而使我国发表在高水平期刊上的论文数量也呈增长趋势。基于此，各高校以及科研机构纷纷采用SCI论文收录数量、引用情况、影响因子等数据作为科研评估的定量指标。我国申报国家自然科学基金、两院院士、长江学者、实验室评估以及科研人才引进等均需提交第三方检索机构出具的SCI论文评估证明。可以说SCI论文相关指标在我国评估科研水平中已不可或缺^[3]。

在国外，SCI数据也广泛地应用到科研绩效评价工作中。比如美国国家科学基金会每两年一次发表的《科学与工程指标》报告，就引用基于SCI的数据；德国教育科研部每年发表关于德国科技竞争力的年度报告，也是基于SCI数据评估；荷兰莱顿大学的科学技术研究中心在SCI的基础上专门建立了为科学计量学分析评估服务的科技指标数据库；英国苏赛克斯大学科学政策研究所是国际科学计量学研究的重要机构，他们以SCI数据为基础建立了自己的面向评价的数据库^[4]。

但是，目前基于SCI数据进行科研评估的争议颇多。一方面，发表SCI论文被认为是科研水平的一个重要标志，SCI论文作为反映科研成果水平的参照系，有利于科研人员对自身研究价值的充分认识，可以帮助科研管理部门快速了解科研项目的基础研究水平，对科研成果进行有依据的评价。另一方面，SCI在实际评估应用中，被有些科研管理机构过分夸大其价值，将其作为唯一依据，并与职称评定、绩效考核、成果奖励等直接挂钩。在评价指标的制定过程中也不考虑评估对象各方面存在的差异，采取“一刀切”的做法，这样势必出现偏颇，有失公正。同时，SCI数据本身也存在着一些缺陷，影响了评价应用的权威性和客观性。

2 SCI用于科研评估的问题

2.1 学科语种分布不平衡

SCI数据库收录论文涉及的学科范围很广，

达到200多个学科，8000多种期刊，但从论文数量来看，医学、化学、生物、物理学等学科所占比例较大^[5]，主要研究还是侧重于基础科学。因此SCI更适合评价基础研究性工作。在采用SCI进行科学研究水平评估时要正确区分是基础研究，是应用研究，还是开发研究，不能把SCI作为唯一的标准，盲目地应用于所有领域科研水平的评价。

SCI的刊源主要来自于英文语种科技期刊，这就造成了一些高水平的科技期刊由于语言原因未被收录。以我国大陆地区为例：2016年SCI数据库共收录期刊8856种，其收录期刊仅有178种，占比2.29%；中文语种期刊仅有22种^[6]。2016年，发表SCI论文32.4万篇，排名世界第二位，但发表的期刊主体多为国外出版机构^[7]。事实上，我国有一些期刊或论文虽然没有入选SCI，但学术影响力却很高。例如：《地理学报》（*Acta Geographica Sinica*）在2007—2016年被SCI论文引用达481次，高于《*Brazilian Journal of Geology*》（440次）、《*Earth Sciences Research Journal*》（283次）等SCI收录的地理类核心期刊^[8]。

2.2 被引频次和影响因子存在差异性

一篇论文的被引频次以及所在期刊影响因子的高低，可以从一个侧面衡量出论文的质量以及学术影响力。但目前国内有些科研管理部门将作者的科学贡献与期刊的影响因子或论文的被引频次直接挂钩，并将其作为唯一的依据。用一把尺子制定奖励标准。实际上，仅仅把影响因子和被引频次作为学术质量的定量指标是有失公正的。

首先，由于各个学科自身发展以及学科结构的特点，论文被引频次在各学科之间存在较大的差异性，分布极不平衡，不同学科论文间的引用率不具有可比性，即便是在同一学科内，不同专业方向的论文引用情况也有很多差别，不能用统一的标准进行衡量和评价^[9]。

其次，影响因子是评价期刊的工具，影响因子的高低不能代表刊载的每篇论文的质量，与论文本身的质量及其科学价值不完全成正比。不同学科领域的期刊影响因子也没有可比性。在SCI

数据中,医学、生物学期刊的影响因子和引用次数就要明显高于数学、地质学等学科。例如:在2015版《Journal Citation Reports》中地质学期刊最高影响因子为4.548,而肿瘤学期刊最高影响因子达到了137.578。有些国际顶尖学术期刊,除了少数极为出色的论文,大部分论文的引用也是远远低于平均值的。常有小部分论文被引频次非常高,从而导致影响因子在均值计算过程中出现偏差。

2.3 部分机构地址数据存在缺陷

SCI数据库将机构名称的多种变体进行了规范化处理,建立了机构扩展字段,可有效避免评估检索时的错检漏检。但SCI数据库并未对全部机构进行机构扩展规范化标引,只有机构发文数量达到一定量级或是主动向SCI数据库商提交申请后才会有机构扩展记录。

对于重点实验室、高校的二级学院、附属医院等二级机构的名称,SCI数据库并没有做进一步的规范化处理。如“华南肿瘤学国家重点实验室”(State Key Laboratory of Oncology In South China)在SCI的地址字段中就会有多种拼写方法(表1)。因无法准确地筛选出二级机构的数据,对评估带来很多困难,

2.4 作者顺序标引不规范

在应用SCI论文进行科研水平评估时,对于第一作者和通讯作者的相关认定存在不少异议。SCI数据库原则上对于第一作者的标注是遵循原文标注,但通过与原文对比后发现有部分第一作者地址的标注数据存在错误。如:当原文中第一

作者所属机构名称太长,SCI数据加工程序会按一定默认规则自动将地址进行切分为两个或多个机构地址,在此过程中会出现一些切分错误。还有一些论文在切分地址后,仅截取了机构名称的缩写。以上地址数据的加工方式均造成了机构名称的不完整,从而无法纳入评估数据中。

同时,由于SCI数据库对于共同第一作者机构并没有进行标注。在进行SCI论文统计时,该指标无法计入统计。而对于共同通讯作者,2015年以前SCI数据库对该指标也没有进行标注,2016年2月才开始全部收录共同通讯作者的信息。部分论文作者字段中对于作者顺序的上角标注与原文不符,导致在评估时对于第一作者和第一作者单位的认定出现异议。以上由于SCI数据本身的缺陷,造成评估数据的遗漏,无法保证评估结果的准确性。

3 SCI数据用作评估指标的建议

3.1 正视SCI数据的评价功能,完善评估体系

目前,国际上许多国家在科研评估中经常使用SCI数据库作为统计源以定量研究科研能力、科学发展趋势与活动水平。将SCI对论文的被引频次统计以及期刊影响因子等指标引入到学术评价体系中,具有其客观性,也更能与国际学术评价方式接轨。以SCI论文为导向,也可激励科技人员向世界高水平研究看齐。

要合理使用SCI就必须很好地消除其弊端。应当制定有导向性的政策,不断完善科研评价体系。任何单一的量化指标都不是绝对的、完善的,是具有一定的适用范围和应用前提的,正确认识和应用SCI对于提高科学研究水平具有重要意义。在进行科研评估过程中SCI指标不应作为首要甚至是唯一的评价标准,要根据评估对象的实际情况,综合考虑科学研究成果的国家目标和社会价值,要考虑不同学科领域研究成果的影响周期、引用规律的差异性等多方面因素。在建立科研评估体系时,定量的SCI指标分析必须结合完善的同行评议机制或辅助其他计量评估方式^[10]。同时,在评估范围的划定时还应考虑到

表1 SCI实验室名称不规范样例(2011—2015年)

SCI中的实验室名称	数量/篇
State Key Lab Oncol South China	1543
State Key Lab Oncol S China	520
State Key Lab Oncol Southern China	278
State Key Lab Oncol So China	247
Key Lab Oncol S China	6
State Key Lab Oncol Sth China	4
Key Lab Oncol So China	3
Natl Key Lab Oncol S China	2
State Key Laboratoty Oncol South China	1

一些具体的特殊情况，以保证评估数据的公平公正。如：机构合并前的论文产出、实验室前身的论文、合作伙伴机构的论文等是否统一纳入评估范围；评估期内人员变动因素是否考虑等。

3.2 提高科研成果管理意识

在网络信息技术迅速发展的今天，科研人员在研究过程中产生了丰硕的数字化研究成果，不论是科研机构还是论文作者都应该提高自身科研成果的管理意识，对其进行数字化收集和规范化保存，这样才能在科研评估时不出现偏差。但在实际工作中，有部分原文中标注的单位信息甚至作者姓名都存在错误，建议作者在论文出版后应仔细核对相关信息，当论文被收录到SCI数据库中，也可能由于各种原因造成数据错误，针对这种情况作者应及时主动与SCI数据库方联系，申请数据修改委托，以保证自己的科研成果能够有效正确地计入评估统计。

发表论文的单位名称规范标注对于统计评价工作是非常重要的，各类不规范表达会造成评估数据的遗漏。基于此，科研机构在实际工作中也应强化论文的规范化标注，同时建立机构知识库，在帮助机构和研究人员保存并展示其学术成果的同时方便科研评估管理。

3.3 完善SCI数据及评价指标

在科学技术飞速发展的今天，许多学术研究是跨学科存在的，机构与机构、人和人之间的合作更为紧密，对于共同第一作者和共同通讯作者的标注也显得尤为重要。一篇高水平的论文由于共同作者的无法认定，会对一个机构的科研水平评估造成损失。SCI数据库方应对共同作者进行标识，并对旧的数据进行回溯整理，以保证数据的公平性和严谨性。

对于错误数据，SCI数据库方应设立自检监察机制，及时进行错误纠正。在对地址进行切分时，应针对不同期刊的著录方式，进一步细化切分规则，对于异常数据应进行人工干预核检。针对二级机构数据无法准确获取的问题，SCI最好能够提供全称检索、原文检索和缩略检索等多种方式，并对同一机构的缩略方式进行统一处理。

同时，SCI数据库应针对每个细分学科，对评价指标进行归一化设定，使得评价指标更为公平公正。

3.4 规范评估检索流程

目前，国内评估检索机构众多，检索方法和水平各不相同，对于评价指标的理解也不一致，所以最好由被评估单位向第三方评估机构提供真实可靠的数据，以保证科技评估的科学性与准确性^[1]。第三方评估机构对于科研评估的检索委托应规范化、流程化处理。首先，在接受检索评估委托时，应该积极参与到前期的SCI评价指标的制定工作中，探讨指标的合理性以及是否通过检索途径能实现。其次，在检索机构SCI论文产出时，背景信息的调研工作也尤为重要，会涉及机构名称变更、机构合并、多家下设机构等情况，在制定检索式时需要整合所有信息，要考虑到各种可能性。在采集SCI数据时，最好能在较短时间内集中下载，以防止数据库更新带来的前后数据不一致。最后，检索完成后应与被检索机构共同核对检索结果，进行查漏补缺的工作，当SCI检索结果与原文标注信息有误差时，双方应共同制定统一标准进行判定。

4 结语

科学研究是一项创新性的活动，其评价工作也极为复杂，SCI论文也只是反映学术水平的一个侧面，其最基本的功能还是帮助科技人员获取最需要的文献信息，如果盲目或过分依赖SCI简单地对科学研究进行评价，势必会出现评价偏倚、急功近利、重数量轻质量、“唯SCI是从”等问题。任何考核评价都是以促进工作、有效激励为目的，制定了正向引导的政策，建立了多角度全面综合的科研产出的评估方案，增加区域性、学科分类等多维度统筹评价对比机制，不断完善科研评估体系，纠正不科学的导向，及时改进目前SCI数据的不足，只有合理使用SCI才能让评估中涉及的每个问题都有规范化的解决方案。科学在发展，学术评估体系要改进完善，本文对应用SCI数据进行科研评估的分析，是很初步的，

有待进一步探讨。

参考文献

- [1] 马英. 浅谈中国学术界“唯SCI论”与学术自信的缺失[J]. 天津科技, 2014, 41(12): 74-76. DOI: 10.3969/j.issn.1006-8945.2014.12.023.
- [2] 刘秀华, 唐朝枢. 对于正确应用SCI作为学术评估体系的一点看法[J]. 生理科学进展, 2009, 40(1): 41-46.
- [3] 孟凡兴. 中国科研体系现状及未来发展研究: 基于我国SCI论文水平的分析[J]. 科技管理研究, 2011(18): 236-240. DOI: 10.3969/j.issn.1000-7695.2011.18.056.
- [4] 武夷山. 关于SCI用作科研绩效评价工具的一些情况[J]. 中国科技期刊研究, 2002(1): 39-41. DOI: 10.3969/j.issn.1001-7143.2002.01.014.
- [5] 桑莱丝. SCI论文在科研水平评价中的地位和作用[J]. 统计与决策, 2007(15): 59-60. DOI: 10.3969/j.issn.1002-6487.2007.15.030.
- [6] 中国科技期刊的困境与出路[N]. 光明日报, 2018-01-18(13).
- [7] 中国科技论文统计结果2017: 中国国际科技论文产出状况[R]. 北京: 中国科学技术信息研究所, 2017.
- [8] THOMSONREUTERS. Journal Citation Reports[R/OL]. [2018-05-11]. http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports.
- [9] 庞龙, 张培富, 杨立英. 引文分析方法在国内外应用的比较研究[J]. 山西大学学报(哲学社会科学版), 2006(3): 134-137. DOI: 10.3969/j.issn.1000-5935.2006.03.030.
- [10] 仇晓春. 替代计量学在图书馆中的应用研究[J]. 数字图书馆论坛, 2017(2): 44-47. DOI: 10.3772/j.issn.1673-2286.2017.2.008.
- [11] 徐志标. 影响因子分数平均值在SCI论文奖励中的应用[J]. 宁波大学学报(人文科学版), 2015(1): 119-122. DOI: 10.3969/j.issn.1001-5124.2015.01.025.

欢迎订阅《中国科技资源导刊》

《中国科技资源导刊》是中国科学技术信息研究所、南京大学共同主办的科技管理类学术性期刊, 是中国科技核心期刊(中国科技论文统计源期刊)。主要刊登科技资源(尤其是科技物力资源、科技信息资源和科技人力资源)管理领域的学术论文、研究报告、综述评论, 宣传和探讨科技资源管理的战略政策, 探索和揭示科技资源管理领域的基本原理和规律, 展示科技资源建设与服务的实践经验等, 促进我国科技资源管理领域的理论研究与实践管理水平的不断提升, 为科

技资源管理者和研究者提供高水平的学术交流平台。

《中国科技资源导刊》读者对象: 政府、科研院所、高等院校、高新技术企业等广大从事科技资源管理的科技管理人员、科研人员和有关专业的研究生及其他研究人员。

《中国科技资源导刊》是双月刊, 大16开, 112页。国际标准刊号: ISSN 1674-1544, 国内统一刊号: CN 11-5649/F。每本定价15元(全年定价90元)。欲购者请与编辑部联系。

欢迎订阅 欢迎赐稿

联系地址: (100045) 北京市西城区三里河路54号《中国科技资源导刊》编辑部

联系电话: 010-68514086 E-mail: zgkjzydk@istic.ac.cn

网 址: <http://www.zgkjzydk.com.cn>