

我国SCI撤稿论文原因分析

靳彤, 张红伟

(中国农业大学图书馆)

摘要: 近年来我国作者的国际学术刊物撤稿事件频发, 引起社会各界的广泛关注, 给我国的学术声誉和科学研究蒙上了一层阴影。撤稿原因是探析撤稿论文的窗口, 本研究以我国SCI撤稿论文为样本, 全面采集分析我国SCI撤稿论文的撤稿声明原文, 根据撤稿声明中描述的具体撤稿原因构建了我国SCI撤稿论文的原因类目体系, 在此基础上对我国SCI撤稿论文的撤稿原因进行了特征分析, 最后就针对性的解决策略进行了相关讨论, 旨在为科研管理部门和科研人员提供参考。

关键词: 撤稿论文; 撤稿原因; 科研诚信

0 引言

2017年4月20日, Springer出版社旗下的《Tumor Biology》期刊发布了一篇撤稿声明, 宣布将撤销107篇涉嫌同行评审造假的论文, 这些论文均出自中国作者之手。这次撤稿事件引起了整个科学界的轰动, 众多学者对于撤稿事件本身及其所暴露的科研诚信问题展开了广泛讨论。2018年10月24日, 南京大学青年长江学者梁某被质疑多篇论文涉嫌抄袭或一稿多投, 当事人主动发起大规模撤稿而引起关注, 社会各界对此议论纷纷, 南京大学也快速做出回应, 经过调查核实梁某存在学术不端行为和违反师德行为, 依据实情做出7项处分。众多撤稿事件将我国学术界推上风口浪尖, 严重影响了我国学术界的声誉, 在一定程度上也反映出我国在科研诚信建设方面存在的一些问题与不足。

1 相关综述

撤稿论文(又称撤销论文或撤回论文)是指已发表的, 经科学共同体成员举报或作者申请、在调查后发现存在科研不端或科学错误等问题的, 对科学发展具有不良影响的, 同时载文期刊针对该文发布正式撤稿声明(撤销声明)、以告知公众的论文^[1]。撤稿论文信息有广义狭义之分^[2], 狭义的撤稿论文信息仅指被撤销论文本身所具有的信息, 包括文献本体特征信息和被引信息等。广义的撤稿论文信息在狭义撤稿论文信息之外还包括与撤稿论文相对应的撤稿声明信息, 主要包括撤稿主体、撤稿原因、撤稿时间等信息。撤稿本质上是学术界自我净化的一种机制, 将存在错误或缺陷的论文从已经公开出版的论文中剔除出来, 保证已有学术研究基础的纯洁性和可信度, 是学术论文出版过程中的重要环节。

国内外学者选择不同的研究样本对撤稿原因展开了大量的实证分析, 根据已有研究经验来看, 造成撤稿的原因是多样的, 部分撤稿论文中的问题源于无心之失, 但还有不少撤稿论文涉及到科研诚信问题。张晴^[3]通过实证分析CNKI中的撤稿论文, 结合国内相关政策, 将撤稿原因归纳为五类——研究成果发表或出版中的科学不端行为、损害他人著作权、研究计划和实施过程非有意的错误或不足、作者责任撤稿(未说明具体原因)和其他。周志新^[4]将CNKI中撤稿论文的原因分为以下五类——涉及

学术不端、涉及知识产权纠纷、作者要求、内容问题和未注明原因。卜今^[5]以PubMed中我国作者的撤稿论文为样本,将撤稿原因分为以下九类:科技错误、编造数据或怀疑编造数据、重复发表或抄袭、署名或版权问题、窃取别人成果、编辑部或出版商错误、基金问题、不符合杂志标准和没有给出撤稿原因或没有信息。秦慧娟^[6]对SCI撤稿论文的撤稿原因进行了分析,将撤稿原因归纳为诚实的错误、抄袭剽窃、伪造篡改、一稿多投重复发表和署名不当五种类型。Fang等^[7]对生物医学和生命科学的撤稿论文进行实证分析,发现科研不端行为是导致撤稿的主要原因。Ribeiro等^[8]通过实证研究将Retraction Watch中的撤稿原因归纳为四类:错误、科研不端行为、其他、不清楚。Chen等^[9]认为中国大陆学者生物医学论文撤稿的主要原因可以分为:欺诈/涉嫌欺诈(数据制作/伪造),重复发表,抄袭,数据错误(数据分析或处理错误、结果不可重复、数据有效性可疑、数据不可靠),发表错误,作者权益争议(未经合著者同意投稿,未经许可发表,作者身份不一致等),伪造的同行评审过程(潜在同行评审员的捏造细节),和其他(未指明的不当行为或上文未提及的原因)。

已有研究选择不同的数据库,不同撤稿群体对撤稿原因进行了初步的探析,列举了常见的撤稿原因,也有研究初步分析了造成论文撤稿背后的深层次原因,这些研究为我们研究撤稿原因,治理撤稿现象提供了有价值的参考。但当前研究对于撤稿原因的分析还缺乏系统性,存在研究中列举的撤稿原因不够全面,部分撤稿原因分类之间存在交叉重叠,部分撤稿原因的具体表现较为模糊等不足。本研究为了全面系统分析撤稿原因,以我国SCI撤稿论文为样本,根据撤稿论文的具体撤稿原因构建撤稿原因类目体系,并在此基础上对我国SCI撤稿论文展开分析。

2 数据来源及处理方法

2.1 样本选择

本研究选择我国SCI撤稿论文作为研究样本,主要源自于以下几点考虑。

一方面,我国撤稿事件多发,撤稿论文数量庞大,我们急需对我国撤稿论文原因给予研究了解。2017年的107篇撤稿事件轰动整个学术界,但国际期刊大规模撤销我国论文并非没有先例。2015年3月26日,开放获取出版商BioMed Central 宣布因评审过程被干扰而撤稿43篇论文,其中41篇来自中国;2015年8月18日, Springer 宣布撤稿64篇论文,其中61篇来自中国作者。在Web of Science中检索撤稿论文可以发现,中国的撤稿论文总量已经位居世界首位,超过发文总量排名第一的美国,远超日本、德国、印度、伊朗等国家。分析我国的撤稿论文原因,能够有针对性地提出缓解我国撤稿现象的治理策略。

另一方面,SCI论文质量相对较高,同时也是我国科研评价、高校评价的重要指标。首先,从普遍意义上来讲,发表在SCI期刊上的论文被引频次高于非SCI期刊论文,SCI论文的质量在一定程度上高于非SCI论文质量,并且SCI论文的出版标准和要求较高,代表着较高研究水平的论文集合。其次,SCI论文受到的关注度更高、接受的监督更广泛,因而SCI论文中存在的问题更容易被发现,存在问题的论文也更容易被撤回。最后,在我国现行科研评价体系中,SCI论文是重要的量化考核指标,SCI论文的数量往往成为科研人员评奖评优、职称评定的关键。所以选择我国SCI撤稿论文进行研究,更有利于我国高校以及科研管理部门了解我国我国SCI撤稿论文原因,并提出有针对性的解决策略。

2.2 数据来源

2.2.1 确定我国SCI撤稿论文集

Web of Science数据库具有强大的论文索引功能，收录了1900年以来的SCI论文集，并在其文献类型（DOCUMENT TYPES）字段中设置“Retracted Publication”和“Retraction”两种类型来标记撤稿论文相关文献，其中“Retracted Publication”文献类型用来标记已被撤回的撤稿论文原文，“Retraction”文献类型用来标记撤稿论文的撤稿声明原文。

为检索我国SCI撤稿论文，本研究在WoS数据库中选择SCI子集，时间跨度选择所有年份，构造检索式如下：

AD = China AND DT = (Retracted Publication OR Retraction)

最终获取检索结果1519条（检索日期为2018年9月26日），下载全部检索结果的元数据全记录，再进行清洗去重。由于部分撤稿论文仅能检索到原文或者撤稿声明，又有一部分撤稿论文的原文和撤稿声明均能被检索到，检索结果中存在小部分重叠，本研究通过题名及作者信息比对，进行清洗去重，最终得到我国SCI撤稿论文1362篇，构建了我国SCI撤稿论文集。

2.2.2 我国SCI撤稿论文原因获取

国内目前尚未有专门的撤稿论文数据库，国际上较为常用的撤稿论文数据库为Retraction Watch Database，但由于该数据库对撤稿原因的记录不规范，且未提供撤稿声明原文，本研究通过以下几种途径来获取撤稿声明原文。

首先，WoS数据库中可检索到部分撤稿论文的撤稿声明，本研究初次检索到的1519条原始数据中有309条数据为撤稿声明原文，可直接下载全文供本研究使用。

其次，WoS数据库中提供大多数撤稿论文原文链接，可通过此链接进入到撤稿论文出版商或期刊原文页面，撤稿论文原文页面通常附有其撤稿声明原文链接，从而获取撤稿声明原文。

最后，WoS数据库中未提供少部分论文原文链接，可从Google或撤稿论文出版商官方网站中检索撤稿论文题名，进入撤稿论文原文页面，再链接至其撤稿声明页面。

Retraction Note to: Mol Neurobiol (2015) 52:1521–1526

DOI 10.1007/s12035-014-8955-7

This article has been retracted at the request of the Editor-in-Chief and the Publisher per the Committee on Publication Ethics guidelines. The article shows evidence of irregularities in authorship.

During the submission process, there is strong reason to believe that the peer review process was compromised and the authors have plagiarized parts from the following article:

Shunzeng Lv, Congxin Dai, Yuting Liu, Bowen Sun, Ranran Shi, Mingzhi Han, Ruixiang Bian, Renzhi Wang, Cell Surface Protein C23 Affects EGF-EGFR Induced Activation of ERK and PI3K-AKT Pathways, Journal of Molecular Neuroscience, February 2015, Volume 55, Issue 2, pp. 519–524, DOI: [10.1007/s12031-014-0375-7](https://doi.org/10.1007/s12031-014-0375-7); Received: 26 June 2014.

As such, the validity of the content of this article cannot be verified.

图1 Springer出版社旗下期刊撤稿声明原文示例

This article has been retracted: please see Elsevier Policy on Article Withdrawal (<http://www.elsevier.com/locate/withdrawalpolicy>).

This article has been retracted at the request of the Editor in Chief and the Authors due to image irregularities in Figure 3A and Figure 5A which affect the understanding and quality of the article.

The authors would like to extend an apology to the readers of Lung Cancer for submitting this paper and for any inconvenience caused.

图2 Elsevier出版社旗下期刊撤稿声明原文示例

2.2.3 数据处理

本研究首先提取撤稿声明中对撤稿原因的原始表述，通常为“reason”、“due to”、“because”等词语后的表述，作为撤稿原因的原始数据。其次，从撤稿原因原始数据中抽取关键词，记录作原始撤稿原因。第三，对原始撤稿原因进行统计分析，了解整体撤稿原因的整体概况，统一、规范同一撤稿原因的不同表述，完成原始撤稿原因的数据清洗工作。再次，结合已有相关研究和经验，将相似撤稿原因进行聚类，形成撤稿原因二级类目团。最后，对撤稿原因二级类目团进行分析归纳，形成撤稿原因一级类目。需要说明的是，由于撤稿原因数据处理的过程中人工干预工作过多，为保证研究结果的准确性和合理性，本研究对于撤稿原因数据清洗及撤稿原因类目构建的过程进行了多次核对确认。

3 我国SCI撤稿原因探析

3.1 撤稿原因类目体系构建

在对撤稿声明原始数据收集和处理的基础上，本研究结合已有相关研究经验构建了撤稿原因类目体系（表1），较为完整详尽地归纳了撤稿原因的类型，并对各类撤稿原因的具体表现做出了列举或解释，尽可能清晰地呈现出撤稿原因的类型和表现，为我国相关管理部门提供参考。

3.2 我国SCI论文撤稿原因分析

在上文对我国SCI撤稿论文原因构建类目体系的基础上，我们对我国SCI撤稿论文的原因特征进行了分析。

3.2.1 科研不端行为是我国SCI论文撤稿的主要原因，抄袭、同行评审造假、重复出版、不当署名是主要表现。

统计分析我国SCI论文撤稿原因发现，科研不端行为是造成我国SCI论文撤稿的主要原因，占比60.67%。统计我国SCI论文具体撤稿原因（表2）发现，抄袭（303次，占比20.27%）、同行评审造假（250次，占比16.72%）、重复出版（136次，占比9.10%）、不当署名（80次，占比5.35%）是撤稿论文中科研不端行为的主要表现，且同一撤稿论文常常同时具备多项科研不端行为。相比于剽窃、篡改、伪造等需要“加工”的造假行为，抄袭这一简单直接的不端方式占比最高。同行评审造假是我国科研不端行为的一种“特产”，因“107撤稿事件”而闻名，国际学术环境中核实和确认专家身份的难度较大，给心怀不轨者提供了可乘之机。涉及不当署名的撤稿论文中，造假作者常常未经合

表1 常见撤稿原因类目划分体系表

一级类目	二级类目	具体表现
诚实的原因	错误	数据错误、图表错误、内容错误、方法错误、技术错误、设计错误、表述错误、结论错误、引文错误、数据可靠性不确定、结论可靠性不确定、结果无法重复、出版错误、数据解释不足、其他错误
	研究过程不规范	在实验准备、数据处理和分析、写作等阶段存在不规范导致文章内容和结果不准确或错误
科研不端行为	抄袭	抄袭数据、抄袭图表、抄袭内容、自我抄袭、无意抄袭
	剽窃	剽窃数据、剽窃内容、剽窃观点、剽窃算法
	篡改	篡改数据、篡改图表
	伪造	捏造、虚造、伪造数据或图表
	重复出版	内容完全或大部分相同的论文被多次出版在同一期刊或不同期刊
	同行评审造假	通过提供虚假同行评审专家信息、伪造同行评审专家意见等方式
	不当署名	合著者不知情或未经同意出版、合著者身份存疑、虚拟作者机构
其他科研不端行为	违反科研伦理、违反期刊政策、作者权益冲突等其他科研不端行为	
其他	未找到或未说明	未找到撤稿声明或撤稿声明中未表明撤稿原因
	重复出版的撤稿论文	曾经被撤稿的论文未经修改或更正再次被出版
	未经内部审查	论文投稿前未经机构内部审查程序而擅自投稿
	作者原因	未说明具体原因，仅表明因作者的原因而撤稿

著者同意或冒用其他作者身份进行投稿，这一造假行为不易发现，在我国SCI撤稿论文中也占据了较大比重。

3.2.2 因诚实原因撤稿的表现形式繁多，数据错误是主要原因类型

因诚实原因的撤稿虽然不是我国SCI论文被撤稿的主要类型，但也表现在不少撤稿论文中，值得引起我们重视和探讨。一方面，撤稿论文中涉及的诚实的错误表现形式繁多，单就错误这一种原因类型就有14种具体表现。从研究前期的方法选择错误、思路设计错误，到中期的数据错误、图表错误、表述错误、内容错误等，再到后期的结果错误、出版错误等，诚实的错误遍布在整个研究过程中另一方面，数据相关的错误是诚实原因中的主要类型。在样本数据中，仅数据错误就出现了226次，在诚实的原因中几乎占据一半的比重，远高于其他类型的错误。此外，数据错误、数据可靠性不确定、数据解释不足可能会导致结论错误、结果无法重复等错误往往可能连环发生，最终给论文的有效性和可信度带来致命打击。

表2 我国SCI论文撤稿原因统计表

一级类目	二级类目	具体表现	撤稿数量	撤稿总比重
诚实的原因 (479)	错误(478)	数据错误	226	15.12%
		结论错误	65	4.35%
		出版错误	43	2.88%
		结果无法重复	35	2.34%
		数据可靠性不确定	28	1.87%
		结论可靠性不确定	19	1.27%
		方法错误	19	1.27%
		图表错误	13	0.87%
		技术错误	8	0.54%
		设计错误	6	0.40%
		表述错误	6	0.40%
		内容错误	5	0.33%
		其他错误	3	0.20%
	数据解释不足	1	0.07%	
	研究过程不规范(1)	研究过程不规范	1	0.07%
科研不端行为 (907)	抄袭(303)	内容抄袭	203	13.58%
		自我抄袭	42	2.81%
		图表抄袭	35	2.34%
		数据抄袭	23	1.54%
	剽窃(67)	数据剽窃	55	3.68%
		内容剽窃	10	0.67%
		观点剽窃	1	0.07%
		算法剽窃	1	0.07%
	篡改(25)	篡改数据	23	1.54%
		篡改图表	2	0.13%
伪造(14)	伪造数据	14	0.94%	
重复出版(136)	重复出版	136	9.10%	
同行评审造假(250)	同行评审造假	250	16.72%	
	不当署名(80)	合著者不知情或未同意发表	74	4.95%
		虚构隶属机构	3	0.20%
		合著者身份存疑	3	0.20%
	其他科研不端行为(32)	作者权益及冲突	18	1.20%
		违反科研伦理	7	0.47%
		科研不端行为	5	0.33%
违反期刊政策		2	0.13%	
其他原因 (109)	其他(109)	未找到或未说明	104	6.96%
		未经内部审查	2	0.13%
		作者原因	2	0.13%
		重复出版的已被撤回的论文	1	0.07%

4 启示与讨论

撤稿原因是撤稿论文的重要属性信息,探析撤稿原因可以了解撤稿论文的作者是否是主观恶意,也可以了解哪些恶意行为是学术不良作者更多使用的。针对这些原因,以及这些原因出现的频率,我们可以从编辑的角度、科研管理机构的角度,以及论文作者自身的角度展开讨论,以寻找有效的办法尽可能避免诚实性错误论文的产生和发表,尽可能地遏制科研不端性质论文的发表,即使发表了也能够尽早地发现并撤销它,尽可能降低其对后续科学研究的不良影响。这样我们能够为期刊编辑加强论文审核指明着力点,也可以为高校、科技部等科研管理部门加强科研诚信教育提供重点教育的方向和内容,还可以针对撤稿作者的科研行为习惯为其他科研工作者提供规范指导等,从而最终为缓解撤稿现象、治理科研不端行为等提供借鉴。

4.1 期刊编辑部加强论文审核,尤其是对撤稿高频原因加以重视

论文从出版到被撤回一般需要经历一段时间,这段时间越长,这些存在问题的撤稿论文对科学界的不良影响就越大。期刊编辑作为论文出版的重要把关人,必须加强论文审核,优化出版论文的质量,尽可能提前将存在问题的论文拒之门外。要做到这一点并不是一件容易的事情。一方面要加强论文审查力度。从常见的撤稿原因出发来审核论文中存在的问题,可通过加强论文查重工作、增加作者身份核实环节来筛查抄袭、不当署名、同行评审造假的论文,尤其是要重点从撤稿高频原因的角度对投稿论文进行审核。另一方面要优化审稿专家匹配程序。涉及内容方法错误、伪造篡改数据等原因的论文不易被外行察觉,但经验丰富的同行专家可能从中发现端倪。选择适合的专家开展审稿工作,同时在评审过程中可要求科研人员提供实验记录来证明数据的可靠性,尽可能将存在明显错误的论文和造假的论文提前拒之门外。

4.2 高校等科研管理部门对抄袭、同行评审造假、重复出版和不当署名行为加强教育

本研究发现,科研不端行为是造成我国SCI论文被撤稿的主要原因,尤其是抄袭、同行评审造假、重复出版、不当署名等科研不端行为表现较为突出。已有撤稿论文中因科研不端行为而被撤稿的论文占据相当大一部分,归根结底还是在于科研人员的科研诚信意识淡薄,对学术规范不了解或不遵循。科研诚信教育告诉科研人员科研诚信的重要性,学术规范教育告诉科研人员如何落实科研诚信,在加强思想认识的同时也要注重方法论的指导,针对不同原因的撤稿行为开展相应的教育,如对于抄袭、剽窃等行为可加强引用文献相关标准的教育,对于因研究过程不规范的行为可加强实验操作和写作规范教育,对于不当署名的行为可加强知识产权和署名规则教育。高校和科研院所通过加强科研诚信和学术规范教育,引导科研人员树立正确的科研诚信意识,养成良好的科研行为习惯。从而减少科研不端事件的发生,降低撤稿论文数量。

4.3 科研人员自身要提高学术规范意识,不要触碰学术红线

归根结底,科研人员才是学术研究的重要主体,科研人员真正地重视科研诚信问题、加强学术规范意识才有可能从根源上缓解撤稿现象和学术不端行为。从意识层面上来说,科研人员认识到科研诚信和学术规范的重要性,才会保持着对科学研究的敬畏之心,不会因为盲目逐利去铤而走险,触碰学术和做人的底线。从实际操作上来说,科研人员认识到科研诚信和学术规范的重要性,才会可能地规范实验操作从而减少实验误差,在实验过程中做好详尽数据记录以备自查,以提高研究结果的可靠性和准确性;才会尽可能完善研究设计和内容,规范语言表达和格式规范,以提高研究结果的有效性和

真实性。

参考文献

- [1] 范少萍, 张志强. 科学文献数据库中撤销论文分布规律研究——以Web of Science 数据库为例. 情报学报[J], 2014, 33 (4): 375-387.
- [2] 张春博, 丁堃, 刘则渊. 撤销论文信息计量的研究综述. 图书情报知识[J], 2018 (04): 53-66.
- [3] 张晴, 姚长青, 潘云涛, 等. 中文学术期刊撤销论文研究. 中国科技期刊研究[J], 2014, 25 (05): 611-615.
- [4] 周志新. 中文科技期刊被撤销论文特征分析及启示. 中国科技期刊研究[J], 2017, 28 (11): 1065-1070.
- [5] 卜今. 生物医学领域中国作者发表文献的撤稿分析. 编辑学报[J], 2013, 25 (06): 571-573.
- [6] 秦慧娟. 基于SCI的论文撤销研究[D]. 上海: 华东师范大学, 2013.
- [7] Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA[J], 2012, 109 (42): 17028-17033.
- [8] Ribeiro M, Vasconcelos S. Retractions covered by Retraction Watch in the 2013-2015 period: prevalence for the most productive countries. SCIENTOMETRICS, [J] 2018, 114 (2): 719-734.
- [9] Chen W, Xing Q R, Wang H, et al. Retracted publications in the biomedical literature with authors from mainland China. SCIENTOMETRICS[J], 2018, 114 (1): 217-227.