

中国知网数据库中高被引文献与高下载文献类型分析 ——以医药卫生科技类文献为例

王 丽

《中国医药导报》杂志社,100025,北京

摘要 在中国知网数据库中检索医药卫生类2011年1月1日至2014年8月31日下载与被引频次排位前100名的文章并对其进行分析,发现下载排名前100篇文章中这2类合计占56.7%,被引排名前100篇文章中述评类和规范类合计占55.2%,两者重叠的23篇中,规范类和述评类合计占69.6%。认为规范类和述评类文献具有较高的权威性、参考价值大,是能够产生较高下载和较高被引的文献类型,应是期刊组稿的重要方向之一。观察下载次数与被引次数发现,两者没有明显的相关性,无规律可循。认为单纯以一两个数据库的下载、被引数据来做定量分析是不科学的,仅可作为定性参考之用。在研究的100篇文献中,有4篇为一稿多发,均属于规范类文章。认为这类文章承担着行业指导性任务,传播理应更广泛,故应与一般意义上的一稿多投、一稿多发的学术不端行为相区别。

关键词 下载次数;被引次数;文献类型;医药卫生

Analysis of types of highly cited and highly downloaded literatures in CNKI database: taking medicine and health care technology literature for example//WANG Li

Abstract The top 100 medical technology articles in download and citation are searched and analyzed, published from 1 January 2011 to 31 August 2014 in CNKI database. It shows that, in the top downloaded 100 articles, review articles and standard type articles together account for 56.7% of the total articles, while in the top cited 100 articles, these two types of articles account for 55.2%. In the 23 overlapping articles, review articles and standard type articles together amount to 69.6%. The literatures of

standard articles and review articles are of high authority and high reference value, which consequently result in high downloads and citations. We think that it would be an important direction for journal soliciting contributions. There is no significant relationship or any rule between the download and citation times. We have to admit that depending merely on one or two databases to analyze the download and citation is not scientific, but can be used as a qualitative reference. Among the 100 articles, there are 4 articles as multiple-publication, all belonging to standard type articles. We think that, this kind of articles play an important role in industry guidance and ought to be widespread to a broader range; therefore, it is different from academic misconduct such as duplicate submission or multiple-publication in general.

Keywords download times; citation times; document type; medicine and health care

Aauthor's address China Medical Herald, 100025, Beijing, China

文章的被引频次是其学术质量以及学术影响力的重要评价指标,由于大多数情况下文献的被引用次数与该文献质量高度正相关,使得引文分析作为科学评价的方法具有一定的可行性,但是作者引用的文献往往仅占其在研究工作中所阅读过的文献的一部分,那么其中未被引用的文献的价值该如何去体现,下载次数是一个日渐公认的评估指标^[1]。从某种意义上说,

5 参考文献

[1] 任胜利,程维红. 2003—2007年中外科技期刊载文与被引的趋势分析[J]. 编辑学报,2009,21(5):468

[2] Journal Citation Reports[EB/OL]. [2014-10-25]. <http://www.thomsonreuters.com>

[3] 中国科学技术信息研究所. 2008—2013年度中国科技论文统计与分析:年度研究报告[M]. 北京:中国科学技术文献出版社,2014

[4] 张玉华,潘云涛,马峥. 科技期刊自引情况的国际对比研究[J]. 编辑学报,2005,17(1):74

[5] 莫京,任胜利. 国内外科技期刊的期刊自引对比分析[J]. 中国科技期刊研究,2005,16(5):655

[6] McVeigh M E. Journal self-citation in the Journal Citation Reports: Science Edition 2002: a citation study from The

Thomson Corporation[EB/OL]. (2009-04-20) [2014-10-25]. <http://www.thomsonisi.com/media/presentrep/es2sayspdf/selfcitationsinjr.pdf>

[7] 宁笔. 中国科学院SCI收录期刊点评[R]//中国科学院自然科学期刊编辑研究会第24次学术研讨会报告. 长沙:中国科学院自然科学期刊编辑研究会,2014-10-15

[8] 刘冰,游苏宁. 理性看待评价指标:优质发展科技期刊[J]. 中国科技期刊研究,2011,22(5):663

[9] 中国科技论文统计与分析课题:2012年中国科技论文统计与分析简报[J]. 中国科技期刊研究,2014,25(1):27

[10] 马峥,俞征鹿,潘云涛. 中国英文科技期刊出版情况的统计分析[J]. 中国科技期刊研究,2014,25(10):1277

[11] 任胜利,宁笔,程维红. 2013年度中国SCI期刊引证指标评析[J]. 科技导报,2014,32(28/29):117

(2014-11-01收稿;2014-12-12修回)

下载频次往往与该文献的被阅读次数相对应^[2],可以更直接地显示期刊文献被读者使用的情况,从而避免一些人为因素(如引文行为和引文动机)对引文分析数据的干扰,而且,从论文上网到被读者下载,其间相对较短的延时使得下载次数能够作为论文价值的快速反映。普遍认为,文章被阅读的次数越多,被引用的可能性越大。那么文章的被引频次与下载次数相关性如何,哪类文献易被引用?笔者利用CNKI检索数据库进行资料收集,对上述问题进行了分析,有助于为期刊组稿工作提供参考。

1 数据与方法

1.1 数据来源 以中国知网(CNKI)数据库为统计源,在“高级检索”页面下,以医药卫生科技为“选择学科领域”,“检索控制条件”设为发表日期从2011-01-01到2014-08-31,检索日期是2014-09-30。

1.2 处理方法 按上述条件,将检索出的文献分别按“被引”和“下载”从高到低的顺序排列,分析排在前100位的文献类型,并分析被引次数与下载次数之间的相关性。本研究的文献类型依据业内习惯设置为规范类(如指南、专家共识、标准及其解读),述评类(如综述,学术争鸣,新药、新技术、新方法的介绍),产业研究类(如医药卫生相关政策的解读、研究,医药市场的分析,医院药企的管理等),基础研究类(医药技术、药物的临床前研究),经验交流类(业务实践中的总结思考),调查统计类(流行病学调查、全国性或省级范围的情况调查等)。为避免文献类型归类不准确,特要求全部被研究文献由一名对医药卫生期刊非常熟悉的人员进行归类,若从题名难以判断的,进一步查阅其摘要以助归类。

2 结果

2.1 下载和被引排名前100篇文章的分析 下载排名前100篇文章中有1篇为杂志稿约,2篇为在他刊重复发表:《中国高血压防治指南2010》分别发表于《中华高血压杂志》(2011-08-15)、《中国医学前沿杂志》(电子版)(2011-10-20),《中国急性缺血性脑卒中诊治指南2010》分别发表于《中国临床医生》(2011-03-05)、《中国全科医学》(2011-12-15)。本研究数据分析时排除该3篇。对其按文献类型分,述评类所占比例最高(39.2%),其次是产业研究类(22.7%)和规范类(17.5%)。被引排名前100篇文章中有4篇为在他刊重复发表:《慢性乙型肝炎防治指南(2010年版)》分别发表于《中国肝病杂志》(电子版)(2011-03-20)、《中华实验和临床感染病杂志》(电子版)

(2011-02-15),《中国2型糖尿病防治指南(2010年版)》分别发表于《中国糖尿病杂志》(2012-01-20)、《中国医学前沿杂志》(电子版)(2011-12-20),《中国高血压防治指南2010》分别发表于《中华高血压杂志》(2011-08-15)、《中国医学前沿杂志》(电子版)(2011-10-20),《中国急性缺血性脑卒中诊治指南2010》分别发表于《中国临床医生》(2011-03-05)、《中国全科医学》(2011-12-15)。本研究数据分析时排除该4篇。对其按文献类型分,述评类所占比例最高(36.4%),其次为调查统计类(27.1%)和规范类(18.8%)。见表1。

表1 下载和被引排名前100位的文献类型分析[n(%)]

文献类型	下载排名前100位	被引排名前100位
述评	38(39.2)	35(36.4)
产业研究	22(22.7)	4(4.2)
规范	17(17.5)	18(18.8)
基础研究	10(10.3)	1(1.0)
调查统计	8(8.2)	26(27.1)
经验交流	2(2.1)	12(12.5)
合计	97(100.0)	96(100.0)

高下载、高被引的文献类型前3位中,同时出现的是述评类和规范类,说明这2类是最容易产生高下载、高被引的文献类型。高下载中2类合计占56.7%(55/97),高被引中2类合计占55.2%(53/96)。

2.2 同时具备高下载与高引用特征的文献分析 下载排名前100篇和被引排名前100篇文章中,同时出现的有23篇(已排除2篇重复发表的),其中,规范类和述评类合计占69.6%,见表2;但这些文章的下载次数与其被引次数间没有明显的相关性,变化趋势也无明显规律,下载次数多的文献,其被引次数不一定更多。

3 讨论

用下载量来衡量期刊的质量,已得到相当多的期刊界同行的认同,据统计,已达到与影响因子相当的效果^[3]。被引用的文章只是说明此文章所起的部分作用,有些人下载文章只是为了学习,并不发表文章,这和以上所说的“实用性”意思相近。该观点还认为,引文量高,与引用的作者的写作条件和习惯有关,可能作者手头恰好有某本期刊,于是引用了其中的相关文献,这种引用就没有下载这个过程。还有一种情况是,高下载有可能是由于文章的题名和摘要吸引人,但内容并无多少可借鉴之处,或者可能是作者由于种种原因故意不引或引其他不相关文献,这就涉及引文行为和引文动机,引用除了自引和课题组引,还存在否定引用,而且很多引用并无实际价值,情况比较复杂。

表2 下载前100名与被引前100名重叠的文献

序号	文献题名	来源期刊	文献类型	下载次数	被引次数
1	2011年中国艾滋病疫情估计	中国艾滋病性病	调查统计	180	2 647
2	中国急性缺血性脑卒中诊治指南2010	中国全科医学	规范	79	2 065
3	万古霉素临床应用中国专家共识(2011版)	中国新药与临床杂志	规范	89	2 029
4	中国高血压防治指南2010	中国医学前沿杂志(电子版)	规范	38 253	685
5	2010年中国CHINET细菌耐药性监测	中国感染与化疗杂志	调查统计	2 675	263
6	中国肿瘤登记地区2008年恶性肿瘤发病和死亡分析	中国肿瘤	调查统计	3 537	236
7	中国肿瘤登记地区2007年肿瘤发病和死亡分析	中国肿瘤	调查统计	2 131	231
8	我国恶性肿瘤发病现状及趋势	中国肿瘤	述评	4 108	127
9	中国2型糖尿病防治指南(2010年版)	中国糖尿病杂志	规范	12 225	124
10	原发性肝癌诊疗规范(2011年版)	临床肿瘤学杂志	规范	2 880	115
11	糖尿病患者自我管理现状及影响因素分析	中华护理杂志	调查统计	4 611	101
12	中国2009年恶性肿瘤发病和死亡分析	中国肿瘤	调查统计	2 719	95
13	中国偏头痛诊断治疗指南	中国疼痛医学杂志	规范	3 160	89
14	2010年全国第5次结核病流行病学抽样调查报告	中国防痨杂志	调查统计	2 249	88
15	黄芪主要活性成分的药理作用	时珍国医国药	述评	2 825	88
16	大黄化学成分与药理作用研究新进展	中国新药杂志	述评	3 464	71
17	慢性乙型肝炎防治指南(2010年版)	临床肝胆病杂志	规范	2 998	71
18	中国鲍曼不动杆菌感染诊治与防控专家共识	中国医药科学	规范	2 290	63
19	植物多糖的研究及应用进展	中国农学通报	述评	2 045	62
20	新型镇静镇痛药——右美托咪定	中国新药与临床杂志	述评	2 233	58
21	栀子化学成分及其药理作用研究进展	中国新药杂志	述评	2 238	55
22	黄芪化学成分及药理作用研究进展	中国疗养医学	述评	2 069	54
23	中药鉴定学新技术新方法研究进展	中国中药杂志	述评	2 589	53

检索当日,我国最大的医学数据库是中国知网和万方数据,在各自的官网上查询,前者收录“医药卫生科技类”(导航条上的显示分类)期刊7 260种,后者收录“医药卫生”(导航条上的显示分类)期刊1 222种,故本文选择收录期刊更加广泛的中国知网为检索数据库,以在中国知网检索得到的数据为统计基础,对其2011年1月1日—2014年8月31日的下载前100位和被引前100位文章进行统计分析。结果显示,规范类、述评类和调查统计类文献的下载和被引均很高,这类文献具有较高的权威性和参考价值,应是各期刊组稿的重点方向。一般来说,论文的下载量和被引量应呈现正相关关系,被下载了才有被引用的可能,即一篇文章的下载次数应高于被引次数;但表3显示,23篇文献中有3篇的下载次数远远低于被引次数。这可能与数据库的统计数据不准确有关,如读者可能是在其他数据库下载阅读,但产生的引用却被CNKI统计了,也可能与读者的不良引用行为(如间接引用)有关^[4]。而且,目前国内几大数据库的竞争也比较激烈,各自签署“独家”的期刊不同,如中华医学会出版的系列杂志于2008年2月底与万方数据库签署了独家合作协议,在中国知网数据库中无法查阅其系列杂志2007年之后的文献。所以,单纯以一两个数据库的下载、被引数据来做定量分析是不科学的,仅可作为定性参考之用。

本研究发现,《慢性乙型肝炎防治指南(2010年版)》分别发表于《中国肝脏病杂志》(电子版)、《中华实验和临床感染病杂志》(电子版)、《中国2型糖尿病防治指南(2010年版)》分别发表于《中国糖尿病杂志》、《中国医学前沿杂志》(电子版)、《中国高血压防治指南2010》分别发表于《中华高血压杂志》、《中国医学前沿杂志》(电子版)、《中国急性缺血性脑卒中诊治指南2010》分别发表于《中国临床医生》《中国全科医学》。一稿多投、一稿多发一直是期刊界深恶痛绝的学术不端现象,主要是会浪费编辑人员劳动和版面资源,也可能引发版权归属的纠纷^[5-7];然而,也有学者从法律角度提出,一稿多投有其合理性及合法性^[8-9]。关于一稿多投、一稿多发的争议越来越多。

一般来讲,下列情况的重复投稿或发表不在一稿多投(发)的范畴:在专业学术会议上做过口头报告,或者以摘要或会议板报形式报道过的研究结果,但不包括以会议文集或类似出版物形式公开发表过的全文;对首次发表的内容充实了50%以上新数据的学术论文;有关学术会议或科学发现的新闻报道,但此类报道不应通过附加更多的资料或图表而使内容描述过于详尽等^[10]。

就本研究中出现的4篇高被引、高下载的一稿多发现象而言,笔者认为与一般意义上的一稿多发不同。