

论文下载量与被引量相关关系的元分析

谢娟, 龚凯乐, 成颖, 戚尔鹏

(南京大学信息管理学院, 南京 210023)

摘要 被引量是广泛研究与应用的文献计量指标。随着出版数字化, 论文下载量指标也受到学界越来越多的关注。目前, 学界就下载量与被引量的相关关系暂无一致结论。据此, 本文利用元分析方法探讨下载量与被引量相关关系的综合效应。通过系统地文献检索与筛选, 获得了 29 项研究和 85 个独立效应量, 涉及样本论文共 115512 篇。通过随机效应模型的元分析发现: 下载量与被引量呈正性的强相关关系 ($r=0.592$)。调节效应分析显示: 论文质量、引证时间窗、下载时间窗及下载数据源对二者的相关关系存在调节效应, 其中, 论文质量的调节效应最为明显。研究结果显示下载量实时、易获取的特点, 使其可作为引用指标的有效补充; 二者的强相关关系提示下载量可以作为科研评价指标之一。

关键词 下载量; 被引量; 元分析; 相关性; 调节效应

Meta-analysis of the Correlation between Downloads and Citations at Paper Level

Xie Juan, Gong Kaile, Cheng Ying and Qi Erpeng

(School of Information Management, Nanjing University, Nanjing 210023)

Abstract: Citation data is a widely researched and used bibliometric indicator. With the development of digital publishing, download data for papers has received increasing attention from academic communities. Currently, the relationship between downloads and citations remains controversial. Therefore, a meta-analysis is used to explore the combined effects of the two, where 85 effect sizes in 29 sample studies are merged, involving 115,512 papers. It is shown that the correlation between downloads and citations is strong ($r=0.592$). The quality of research papers exhibits a strong moderating effect, while citation windows, download windows, and download sources also exert some extent of moderation. It is revealed that download data is conclusively proved to provide an index for research evaluation, owing to its strong correlation with citations. Furthermore, with its advantages of instantaneity and accessibility, it also acts as an effective complement to citations.

Key words: downloads; citations; meta-analysis; correlation; moderating effect

1 引言

引证行为是学术交流、知识共享以及信息融合的客观体现。基于引证行为的引文分析, 已经发展出一套相对完备的方法体系, 成为图书情报学科极

具专业特色的方法论输出^[1]。作为引文分析的基石, 被引量是衡量学术影响力的重要指标。与此同时, 由于被引量存在的固有缺陷^[2], 学界对替代指标的探索从未间断。

目前, 绝大多数学术期刊都可以在线获取, 这

收稿日期: 2017-06-25; 修回日期: 2017-12-10

基金项目: 国家社会科学基金“施引者引用意向与文献计量视角的学术论文被引影响因素研究”(17BTQ014)。

作者简介: 谢娟, 女, 1995 年生, 博士研究生, 主要研究方向为信息计量; 龚凯乐, 男, 1993 年生, 博士研究生, 主要研究方向为信息分析。成颖, 男, 1971 年生, 博士, 教授, 博士生导师, 主要研究方向为用户信息行为、信息检索, E-mail: chengy@nju.edu.cn; 戚尔鹏, 男, 1992 年生, 硕士研究生, 主要研究方向为数据挖掘、机器学习。

使得国内外许多学术数据库(如 Web of Science、ScienceDirect 以及 CNKI 等)能够以更为及时、便捷的渠道记录、存储并提供以浏览量和下载量为代表的使用指标(Usage Metrics)信息^[3]。本质上,使用数据反映了读者对文献的关注度,而其中的部分读者可转化为施引者。据此可以认为,使用数据相较于引用数据具有更加广泛的影响力^[4]。

已有学者对下载量的分布规律^[5-7]、峰值时间^[8-11]及老化规律^[12-14]进行了揭示。多项研究探讨了下载量与被引量之间的相关性:部分研究中二者呈强相关关系,进而学者们探索了通过下载量预测被引量的可行性^[15-20];也有学者观察到两者间的相关关系较弱^[12,21-23],甚至也有负相关关系的报告^[24-25]。Moed 等^[11]以内省的方式阐述了 10 种可能影响实证研究中被引量与下载量相关关系的原因,Wang 等^[5]认为除指标性质和技术缺陷外,样本选择可能也是造成实验结果偏差的主要因素。

相关研究中下载量与被引量相关关系的较大差异给使用指标的开发和应用造成了较大障碍。基于此,本文采用元分析方法,统合 29 项实证研究中的 85 个效应量数据,探讨下载量与被引量的综合效应以及影响两者关系的调节变量,以消解研究间的不一致性,为科研评价替代指标的探索提供新思路。

2 文献回顾

2.1 下载量

下载量指标是出版数字化的直接产物。作为科学计量学的新秀——使用指标之一,下载量记录了论文在期刊网站或出版商中被保存、下载的次数^[26],它与直接反映论文被引的引用指标有异,也有别于以社交媒体影响力为基础的替代计量学指标^[27]。

关于下载量分布规律、峰值时间及老化规律的研究,引起了学者的广泛兴趣。Wang 等^[5]研究发现,使用数据呈现明显的偏态分布,幂律函数可以较好地拟合其分布曲线,该研究表明,仅少数论文被大量阅读和下载,说明论文的使用也具有马太效应。方红玲^[8]和 Schlögl 等^[28]分别对国内外论文下载量达到峰值的时间进行探讨,得到了一致结论,即论文在发表后 1~2 年获得的下载最多,且数年后下载量再次增加,提示论文被引的增多也会一定程度上影响下载。Moed^[12]对单本期刊的下载量进行了历时和共时分析,发现该刊论文的下载量在 6 个月之内就能达到总下载量的 40%,且该刊论文的下载半衰期

为 12.6 个月,明显短于被引半衰期;Schlögl 等^[29]的研究也得出了相同的结论。Wan 等^[30]提出了下载即年指标(Download immediacy index),定义为期刊论文出版一年内的下载量与该刊当年发表的论文总量之比,该研究显示 DII 可以作为独立指标使用,且对期刊 h 指数等指标具有一定的预测价值。

与此同时,对下载量影响因素的探讨也应运而生。Botting 等^[18]通过方差分析发现社交媒体暴露会影响论文的下载量。Shuai 等^[31]、Hajjem 等^[32]和 Wang 等^[33]分析了开放存取对论文下载量的影响。Jamali 等^[34]通过比较不同标题的论文发现提问式标题和较短标题的论文被下载的次数更多。也有学者就语种展开了探讨^[35],发现英文文献具有明显的下载优势,表明下载量受期刊、论文传播情况的影响。

2.2 下载量与被引量的相关关系

被引量作为目前科研评价的重要指标,反映了基于文献的学术交流过程,其仅适用于以“publish or perish”为宗的部分学术团体,但随着网络的普及,许多学科的学术影响力已经扩散到社交媒体、ResearchGate 等非正式学术交流平台^[27]。使用指标和替代计量学指标的出现弥补了传统学术交流模式下评价不全面、不及时的缺陷,而下载量指标较之替代计量学指标的稳定性^[27],使其备受推崇。于是,下载量与被引量的相关关系研究引起了学界的关注。

大量研究从论文层面对下载量与被引量的相关关系进行了探讨。Schlögl 等^[28]研究发现,图书情报学领域论文的下载次数与被引次数相关程度较高($r_s=0.770$ 和 $r_s=0.760$),而 Moed^[12]以 *Tetrahedron Letters* 为例,发现该刊论文的下载量与被引量间关系较弱($r_s=0.220$)。Botting 等^[18]试图根据下载量预测论文的被引量,发现论文发表年内所获得的下载量可以预测其 3 年后的被引情况($r=0.450$)。牛昱昕等^[36]的研究则表明,单篇论文的下载频次与被引频次间的相关性不显著。

也有学者从期刊层面进行了相关研究。Bollen 等^[37]仿照 ISI 期刊影响因子的定义,率先提出使用影响因子指标。Gorraiz 等^[38]的数据选取了 4 个领域的 362 本期刊,分别比较了期刊使用因子(JUF)、Garfield's 影响因子(GIF)和总影响因子(TIF)之间的关系。庞景安^[39]探讨了中文科技期刊被引与下载指标间的关系,发现总下载量与总被引量的相关系数为 0.545,而影响因子前 100 名的期刊总下载量与总被引量的相关性则更强($r=0.659$)。

少数研究以个人、机构为单位, 探讨下载量与被引量的关系。Kurtz 等^[40]采集美国天文学领域中 922 位博士的论文数据, 试图明确论文下载量与被引量的相关关系以预测个人未来成就。Boukacem-Zeghmouri 等^[41]则从机构层面展开了类似研究。

可以看出, 上述研究以单篇论文的下载量与被引量为基础数据, 汇总并衍生出期刊、个人、机构等层面的下载与被引间的相关性分析。各研究结论的明显差异为本文从单篇论文层面统合分析下载量与被引量的相关关系提供了依据。

2.3 下载量与被引量相关关系的调节变量

学科。Wan 等^[30]发现, 农学论文的被引量与下载量具有较强的相关关系 ($r=0.635$); Subotic 等^[23]的研究数据取自于心理学的高下载量论文, 结果显示下载量与被引量之间的相关系数仅为 0.220, 作者认为该结果可能是心理学的学科性质造成的。

期刊类型。Nieder 等^[22]根据 BMC 期刊网站中 5 本肿瘤学期刊的论文数据, 发现下载与被引之间关系较弱 ($r=0.230$), 而 Liu 等^[42]以 5 本中国眼科学期刊发表的论文为研究对象, 发现被引量与下载量的相关系数为 0.491, 呈中度相关。Schloegl 等^[43]的研究则发现 SCI 一区的药学期刊 *Drug Discovery Today* 中论文下载量与被引量呈强相关关系 ($r=0.560\sim0.770$)。

文献类型。Ketcham^[44]探讨了 Lab. Invest 期刊网站中论文的下载与被引情况, 发现该刊中被下载最多的是技术报告而非综述, 且技术报告中两者的相关性较综述低。Moed^[12]认为其研究中被引量与下载量相关性较低的原因与期刊 *Tetrahedron Letters* 上发表的短讯 (short communications) 较多有关。牛昱昕等^[36]的研究显示应用型、综述型论文更容易出现“高下载低引用”现象。

时间。Schloegl 等^[29]比较了论文发表后 1-5 年下载量与被引量的关系, 结果发现论文发表 2 年后所获得的下载和被引之间的相关系数为 0.510, 而发表 5 年后的下载量与被引量的相关关系更强 ($r_s=0.630$)。

数据源。Li 等^[6]以 PLoS 网站的论文为研究对象, 发现其被引量与下载量的相关系数为 0.402, 这一结果与 Brody 等^[16]和 Shuai 等^[31]对 arXiv.org 中论文下载与被引的分析结果相近 ($r=0.440$ 和 $r=0.387$)。O'Leary^[45]将单个期刊中下载量最高的 25 篇论文分别与其在 SSCI、Scopus 和 Google 中获得的被引量

进行比较, 结果表明尽管被引数据源不同, 所得相关系数差异不大 ($r=0.839$ 、 0.820 、 0.784)。

3 数据与方法

3.1 检索策略

第一步, 浏览。浏览涉及下载量与被引量研究的重点期刊 (如 *Scientometrics*、*Journal of informetrics*、*Journal of the Association for Information Science and Technology* 以及《情报学报》等) 的目录以及参考文献, 初步确定有关下载量与被引量相关关系研究可能涉及的检索词。

第二步, 检索。于 2017 年 5 月初完成, 检索范围为 Web of Science、Scopus、PubMed、EBSCO LISTA、SpringerLink、Wiley、Emerald 和 CNKI 以及万方数据库。检索词为 (“download*”) AND (“citation*” or “quotation*” or “cited paper*” or “cited document*” or “cited data*”) 及相应中文检索词, 将上述检索词输入主题字段进行检索。文献类型包括期刊论文、会议论文、学位论文等。通过检索结果标题以及摘要的阅读, 没有发现更多的检索词。对检索结果中的相关论文进行全文下载, 获得可能纳入元分析的研究 77 篇。

第三步, 扩展检索。对第二步中检索到的 77 篇文献通过其引用与被引进行扩展检索, 另外获得了 4 篇论文, 合计 81 篇文献。

3.2 纳入排除标准

本文探讨论文下载量与被引量的相关性, 选取相关系数 r 作为效应量 (Effect Size, ES)。由于 Spearman 等级相关系数与 Pearson 相关系数具有相同的抽样误差方差 (Sampling error variance)^[46], 据此, 本元分析将少数报告 Spearman 等级相关系数的研究也予以纳入。部分研究在探讨下载量与被引量的相关关系时, 采用回归分析、参数检验等方法, 鉴于标准化回归系数 β ^[47]、 t 统计量^[48]、 F 统计量^[48]等统计量可以转换为相关系数, 一并纳入。考虑到样本量是元分析研究中权重计算的依据, 据此, 本研究制定了以下 2 条纳入标准:

(1) 报告了论文下载量与被引量相关系数 (包括 Pearson 和 Spearman), 或报告了其他可转换为相关系数的统计量。据此, 排除文献 20 篇。

(2) 报告了样本量。据此, 排除文献 7 篇。

同时, 为排除内容不相关的研究, 本文制定了

以下5条排除标准:

(1) 综述、述评、编者的话等非实证研究。据此,排除文献3篇。

(2) 探讨开放存取、社交媒体曝光、论文刊载顺序、语种等因素分别对下载量和被引量的影响,未涉及二者相关关系的论文。据此,排除文献8篇。

(3) 本文元分析的对象是论文层面的下载量与被引量间相关关系,排除从期刊、个人、机构等集合层面探讨二者关系的研究。据此,排除文献12篇。

(4) 讨论下载量衍生指标的相关研究。据此,排除文献1篇。

(5) 中英文文之外的研究。据此,排除文献1篇。最终共得到29篇元分析样本。

3.3 编码

针对研究问题及原始研究提供的数据,制定数据编码表。编码变量包括:作者及论文发表年份、样本论文发表年份、样本量、样本论文来源学科、样本论文来源期刊、样本论文质量、引用时间窗(Citation windows)、下载时间窗(Download windows)、被引数据源、下载数据源及相关系数。其中:

作者及论文发表年份。指纳入元分析的原始研究的作者及其发表年份,若同一作者同一年份有多篇论文纳入,用小写字母标识;若同一论文中存在多项独立研究,用数字和括号标识,例如,“O’Leary2008a(1)”。

样本论文发表年份。各原始研究中所选样本的发表年份。

样本论文来源学科。各原始研究中所选样本的来源学科,按Web of Science数据库的标准^①分为生命科学与生物医学、自然科学、应用科学、艺术与人文科学、社会科学五类,在meta回归中分别用数字1~5标识。

样本论文来源期刊。各原始研究中所选样本的来源期刊。

样本论文质量。各原始研究中所选样本论文的质量等级,用样本论文来源期刊在发表年份被SCI或SSCI收录的JCR分区标识。若其在发表年份未被SCI或SSCI收录,数年后被收录,按被收录后的分区编号进行标识。未被SCI收录的中外文自然科学

期刊统一编为4级,未被SSCI收录的外文社会科学期刊编为4级,未被SSCI收录的中文社会科学期刊根据南京大学文科业绩点标准^②编码:属于南京大学学科一流期刊目录^③的编为3级,其他CSSCI^④收录期刊编为4级。若所选样本来源于多年份、多期刊,则用各年份期刊分区的秩平均数标识。例如,Schlögl等^[28]从*Journal of Strategic Information Systems*期刊2002-2011年发表的论文中抽取150篇进行研究,这十年间,该刊的JCR分区分别为:1区、3区、3区、3区、2区、3区、2区、1区、1区、2区,于是有 $(1+3+3+3+2+3+2+1+1+2)/10=2.1$,四舍五入后则该研究对应的样本质量编为2。

引用时间窗。各原始研究中获取被引量的年份与所选样本论文的发表年份之差。

下载时间窗。各原始研究中获取下载量的年份与所选样本论文的发表年份之差。

被引数据源。各原始研究获取被引量的平台,分为Web of Science、Scopus、CNKI、期刊网站(如PLoS网站、BMC网站)、搜索引擎及开放存取平台(如arXiv.org、RePEc数字图书馆)六类,在meta回归中分别用数字1~6标识。

下载数据源。各原始研究获取下载量的平台,分为ScienceDirect、CNKI、期刊网站、搜索引擎及开放存取平台五类,在meta回归中分别用数字1~5标识。

若原始研究中未报告相关信息,本文尝试以电子邮件的方式联系作者,若仍无法获取缺失信息,则在相应编码栏中标“—”。上述步骤由两位研究者经过培训后独立完成。遇到编码不一致则通过协商解决。若协商未果,则寻求第三位研究者的意见,以获得一致编码。

3.4 元分析方法

当不同研究的数据来自于同一整体时,理论上其结果应该一致,这时对其进行合并可以获得真实效应的最佳估计,但文献计量学研究通常不满足该假设^[49]。因此,本研究采用随机效应模型,该模型可同时考虑研究内部和研究间的差异,能降低犯I型错误的概率、更好地估计真实效应^[50]。本研究选择相关系数 r 经过fisher转化后的 z 值作为最终效应

① Web of Science (WoS) 学科分类 http://images.webofknowledge.com/WOKRS56B5/help/zh_CN/WOS/hp_subject_area_terms_easca.html

② 南京大学文科业绩点标准 <http://scit.nju.edu.cn/Item/1234.aspx>

③ 南京大学文科一流期刊目录(2017年修订) <http://skch.nju.edu.cn/regulation/1686950903>

④ 中文社会科学引文索引 <http://cssrac.nju.edu.cn/a/zlxz/20160329/2720.html>

量^[51]。对于未直接提供相关系数的研究,按照相应公式进行转换,比如, Jahandideh 等^[17]对 2006 年 ScienceDirect 数据库下载最多的 25 篇和最低的 25 篇论文与其被引量进行了 t 检验,得到统计量 $t=2.212$,根据参考文献[48]提供的公式进行转换,得到相关系数 $r=0.304$ 。在报告结果时,再将 z 转换回相关系数 r 。异质性检验通过 I^2 统计量完成。发表偏倚通过漏斗图、Rosenthal's 失安全系数法和 Egger 回归法进行检验。通过 meta 回归找到可能的调节变量,并进行亚组分析。上述过程均通过 stata14.0 完成。

4 研究结果

4.1 编码结果

经过详细地编码与校验,从 29 个原始研究中提取了数据(附录 1)。纳入研究中含有英文文献 22 篇,中文文献 7 篇;各原始研究的样本量范围为 10~33128 篇、样本论文发表时间跨度为 1994~2014 年。Jamali 等^[34]13 篇研究选择特定时间段内发表的论文,探讨其数年后下载量与被引量的关系,其他研究则采用随机取样。Schloegl 等^[29]等研究中存在多项研究,相应地可获得多个独立的相关系数,最终得到 85 个效应量。

Cohen^[52]提出了判断相关系数 r 强弱的经验准则: r 介于 0.00~0.09 表示基本无相关关系;0.10~0.29 为弱相关;0.30~0.49 为中度相关;0.50~1.00 为强相关。获得的 85 个效应量中, Schlögl(1)等^[28]等 53 项研究发现论文下载量与被引量具有强相关关系;Li 等^[6]等 20 项研究结果显示下载与被引中度相关;Nieder 等^[22]等 6 项研究结果则呈弱相关;丁佐奇等⁽²⁾^[15]和杜秀杰等⁽⁵⁾^[53]未发现两者间存在相关关系;接雅俐等⁽²⁾^[24]等 4 项研究报告了二者间的负相关关系。

4.2 主效应

图 1 为运用元分析对论文下载量与被引量相关系数进行合并的结果。结果显示:85 个效应量最终合并结果的 z 为 0.68 (CI: 0.63~0.72),通过 fisher 转换,得到整体相关系数 $r=0.592$,表明论文下载量与被引量具有正性强相关关系,即论文下载量越多,其被引频次也越高。

4.3 异质性检验

Higgins 等^[54]认为 $I^2=0\%$ 时,表明研究间无异质性; $I^2 \leq 25\%$ 存在轻度异质性; $I^2 \leq 50\%$ 存在中度异质

性; $I^2 \leq 75\%$ 存在高度异质性; $I^2 > 75\%$ 时不适合进行元分析。本研究的 I^2 值为 100.0%,表明研究间异质性较大。与本研究同类型的工作——Bornmann^[49]的研究中 I^2 值也高达 99.9%。为了更好地阐释异质性检验结果,本研究依据 Q 与 P 的计算公式,进行了模拟分析。

Q 统计量:

$$Q = \left(\sum \omega_i ES_i^2 \right) - \frac{\left(\sum \omega_i ES_i \right)^2}{\sum \omega_i} \quad (1)$$

I^2 统计量:

$$I^2 = \begin{cases} \frac{Q - df}{Q} & Q > df \\ 0 & Q \leq df \end{cases} \quad (2)$$

其中, df 为自由度。

考虑到进行元分析的前提是原始研究间的效应量存在显著差异,在模拟分析中,ES 在 0.01~0.99 随机取值。

ω_i 是原始研究中的样本量,在模拟分析中分别取医学、行为科学、心理学、商学以及文献计量学中常见的样本区间,使其在 10~30、10~50、50~100、10~1000、10~30000 内随机取值。

考虑到不同元分析研究中效应量的数量甚少超过 100,同时不少于 2,在模拟分析中,效应量的数量在 2~100 之间随机取值。

模拟结果如表 1 所示,其中第一行表示原始研究中的样本量区间,第一列表示元分析中 I^2 的区间,中间数据表示模拟得到的相应元分析的数量。

根据公式(1), Q 值与样本量 ω_i 以及 ES 存在直接关系,其中 ES 的取值区间较窄,因此, Q 值受 ω_i 的影响更为明显。由表 1 的数据可见,对于行为科学研究中的常见样本量(10~30), I^2 普遍在可接受的范围内,当样本量介于 10~100 时,根据 Higgins 的判断准则,已有 69% 的研究不适宜进行元分析,当样本量介于 50~100 时,已有 97% 的研究不适宜进行元分析,随着样本量的进一步增大, I^2 基本都在 0.9 以上,已不符合 Higgins 准则。据此,可以认为 Higgins 的判断准则更多地适用于行为科学、医学等小样本研究,对于诸如文献计量等大样本研究,Higgins 准则很难适用,这也是 Bornmann^[49]的元分析研究能够发表的原因。

4.4 发表偏倚检验

通过漏斗图(funnel plot)对发表偏倚进行定性

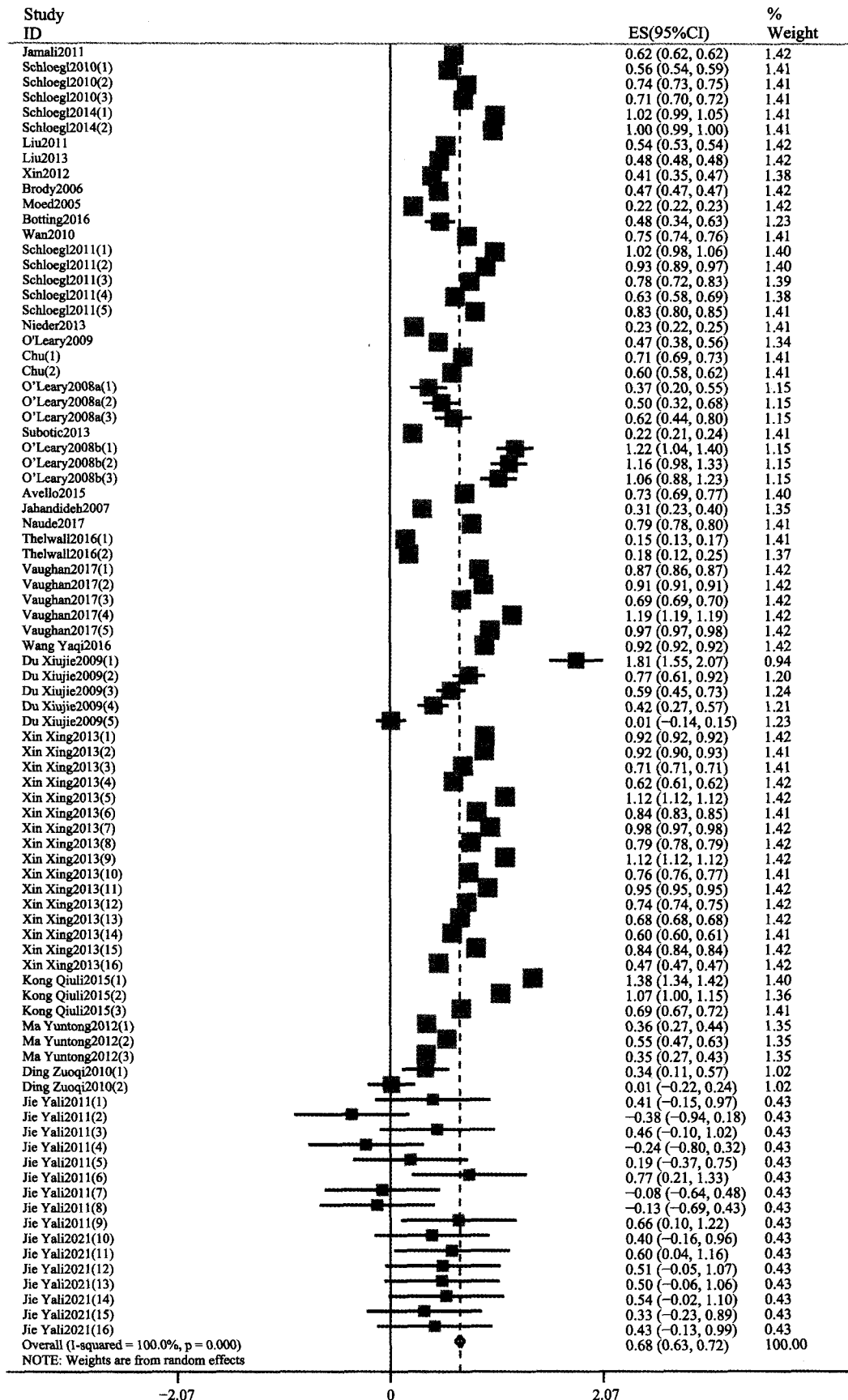


图1 元分析结果

表 1 I^2 模拟结果

| | 10~30 | 10~50 | 10~100 | 50~100 | 10~1000 | 10~30000 |
|-------------|-------|-----------------|-----------------|------------------|----------------|------------------|
| 0~0.25 | 31 | 8 | 0 | 1 | 0 | 0 |
| 0.25~0.5 | 152 | 27 | 0 | 0 | 0 | 0 |
| 0.5~0.75 | 10 | 156 | 70 | 5 | 1 | 0 |
| 0.75~0.99 | | 2 | 134 | 193 | 196 | 196 |
| (I^2 区间) | 0 | (0.76~ 0.79) | (0.75~ 0.87) | (0.755~ 0.95) | (0.9~ 0.98) | (0.99~ 0.999) |

检验。由图 2 可知，各研究的分布并不对称，多数研究分布于上侧、右侧，表明本元分析可能存在一定的发表偏倚。进一步进行定量检验，Egger 回归检验的截距为 99.405 ($p<0.005$)，此时应接受原假设，表明存在发表偏倚。Rosenthal's 失安全系数法计算结果 N 为 4.3×10^8 ，远大于 435 ($5k+10$, k 为效应量个数)，说明至少需要再纳入 4.3×10^8 个效应量才能使元分析结果发生逆转，因此 Rosenthal's 检验表明本元分析不存在发表偏倚。

鉴于上述三种检验的结果不一致，本文通过 Duval 等^[55]提出的剪补法进一步分析发表偏倚是否会对元分析结果造成影响。结果发现，经过 4 次迭代，减掉 17 个效应量，并围绕漏斗图中心在对侧镜像位置补上 17 个类似的研究，最终使得漏斗图两侧的散点大致对称（图 3）。通过分析发现，剪补前后

的合并效应量 z 值分别为 0.68 和 0.53，表明发表偏倚对元分析结果影响不大。因此可以认为，本元分析虽存在轻微的发表偏倚，但结果还是有效的。

4.5 调节效应

meta 回归。采用 meta 回归方法，发现可能的调节变量并从整体上对调节效应进行分析（表 2）。由表 2 可见，因每项纳入研究未涉及所有的调节变量，所以各调节变量与 ES 进行回归分析时的样本数存在差异。根据 t 值与 p 值，可以发现：样本论文质量、引用时间窗、下载时间窗以及下载数据源对下载量与被引量的相关关系具有显著影响；样本来源学科与被引数据源的调节效应不具有统计学显著性。

亚组分析。围绕 meta 回归发现的 4 个具有显著调节效应的变量进行亚组分析（表 3），探讨组间效应量的差异从而深入分析各变量的调节效应。表 3 结果显示：

（1）论文质量。在样本论文质量分组中， I^2 值出现了明显降低，表明论文质量对两者关系的调节作用显著。总体上，meta 回归系数 $\beta=-0.0877$ ，负性的回归系数说明随着论文质量等级的降低，下载量与被引量的相关性随之减弱。分组情况则存在细微的差异：质量等级为 2 级的论文下载量与被引量的相关系数最高，为 0.623，说明该层次论文的质量最为接近；质量等级为 4 级的相关系数最低，为 0.445，说明该组论文的质量差异最大；但质量等级为 1 级的相关系数略低于 2 级，与回归趋势稍有差异，导致该现象的引用行为与引用动机拟另文探究。

（2）引用时间窗和下载时间窗。引用时间窗和下载时间窗对两者的相关性也具有显著影响，且回归系数 β 均大于 0，表明整体上论文发表时间越长，其下载与被引的相关程度越高。各原始研究引用时间窗和下载时间窗的取值除 2 项外一致，故一并分析：表 3 显示，论文发表后，下载量与被引量的相关性逐渐增加，并在发表后第三年达到第一个高峰（ $r=0.675$ ），随后其相关关系略有回调，但在发表五年后，两者的相关关系又继续上升且基本稳定于 0.7 左右。

（3）下载数据源。下载数据源对两者的相关性存在负性的调节效应（ $\beta=-0.0701$, $p=0.038$ ），回归拟合的结果显示从 ScienceDirect、CNKI、期刊网站、搜索引擎到开放存取平台，下载量与被引量的相关关系依次递减。亚组分析表明，下载数据源为 ScienceDirect 和 CNKI 的论文下载量与被引量具有强相关

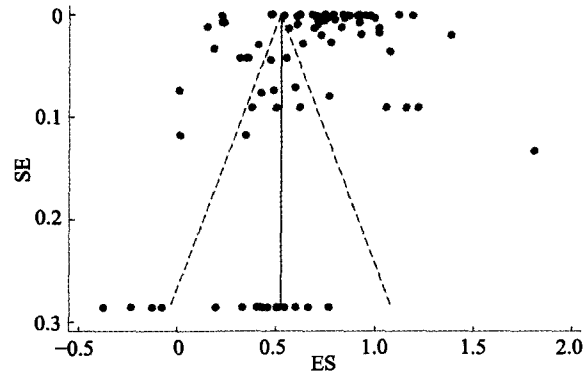


图 2 发表偏倚漏斗图

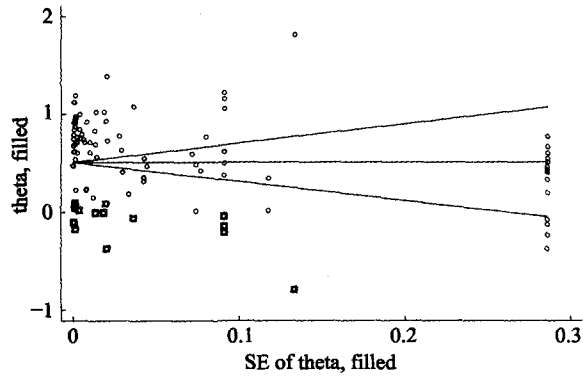


图 3 剪补后的漏斗图

关系，而期刊网站及开放存取平台仅中度相关（ $r=0.478$ 和 $r=0.437$ ）。

表 2 meta 回归结果

| 变量 | 样本数 | 回归系数 β | t | $P> t $ |
|--------|-----|--------------|-------|---------|
| 样本来源学科 | 85 | -0.0345 | -1.43 | 0.158 |
| 样本论文质量 | 49 | -0.0877 | -1.98 | 0.054* |
| 引用时间窗 | 41 | 0.0173 | 1.98 | 0.054* |
| 下载时间窗 | 40 | 0.0152 | 1.79 | 0.082* |
| 被引数据源 | 85 | -0.0021 | -0.06 | 0.955 |
| 下载数据源 | 85 | -0.0701 | -2.11 | 0.038** |

注：**表示 $p<0.05$ ，*表示 $p<0.1$ 。

5 结论与讨论

学界已经形成了多项下载量与被引量的相关关系研究，研究间存在着较大的不一致性， r 介于 $-0.359\sim0.948$ 。本文采用元分析方法，统合分析了下载量与被引量之间的相关关系，并探讨了论文质量等变量的调节效应。

5.1 下载量与被引量的相关关系

主效应分析结果表明：单篇论文下载量与被引量具有较强的正相关关系（ $r=0.592$ ），即论文下载量越大，其被引频次越高。多项研究发现期刊、个人以及机构粒度的二者关系研究也呈强正相关关系^[40-41,56]。

表 3 亚组分析结果

| 调节变量 | 组别 | 效应量个数 | ES | r | 组内 $P/\%$ | Q 统计量 | p |
|---------|----------------|-------|-------|-------|-----------|----------|-------|
| 样本论文质量 | 1 级 | 10 | 0.670 | 0.585 | 99.9 | 12030.48 | 0.000 |
| | 2 级 | 8 | 0.731 | 0.623 | 99.9 | 8149.89 | 0.000 |
| | 3 级 | 6 | 0.614 | 0.547 | 86.2 | 36.32 | 0.000 |
| | 4 级 | 25 | 0.479 | 0.445 | 99.0 | 2386.76 | 0.000 |
| 引用时间窗/年 | 1 | 2 | 0.620 | 0.551 | 99.4 | 166.43 | 0.000 |
| | 2 | 11 | 0.558 | 0.506 | 100.0 | 6.8e+05 | 0.000 |
| | 3 | 3 | 0.820 | 0.675 | 98.9 | 188.71 | 0.000 |
| | 4 | 6 | 0.649 | 0.571 | 99.9 | 3871.49 | 0.000 |
| | 5 | 7 | 0.913 | 0.723 | 100.0 | 73146.20 | 0.000 |
| | 7 | 4 | 0.894 | 0.713 | 100.0 | 4.7e+05 | 0.000 |
| | 12 | 4 | 0.930 | 0.731 | 100.0 | 1.9e+05 | 0.000 |
| | 17 | 4 | 0.792 | 0.660 | 100.0 | 61125.50 | 0.000 |
| 下载时间窗/年 | 1 | 3 | 0.667 | 0.583 | 98.8 | 167.15 | 0.000 |
| | 2 | 9 | 0.574 | 0.518 | 100.0 | 6.8e+05 | 0.000 |
| | 3 | 3 | 0.820 | 0.675 | 98.9 | 188.71 | 0.000 |
| | 4 | 6 | 0.649 | 0.571 | 99.9 | 3871.49 | 0.000 |
| | 5 | 7 | 0.913 | 0.723 | 100.0 | 73146.20 | 0.000 |
| | 7 | 4 | 0.894 | 0.713 | 100.0 | 4.7e+05 | 0.000 |
| | 12 | 4 | 0.930 | 0.731 | 100.0 | 1.9e+05 | 0.000 |
| | 17 | 4 | 0.792 | 0.660 | 100.0 | 61125.50 | 0.000 |
| 下载数据源 | Science Direct | 19 | 0.731 | 0.624 | 100.0 | 46625.00 | 0.000 |
| | CNKI | 53 | 0.710 | 0.611 | 100.0 | 2.1e+06 | 0.000 |
| | 期刊网站 | 5 | 0.520 | 0.478 | 100.0 | 27735.90 | 0.000 |
| | 开放存取 | 8 | 0.468 | 0.437 | 99.6 | 1659.27 | 0.000 |

文献[57]提供了可能的答案：Web2.0 环境下，学术文献的下载与阅读是合理引用的必要前提，只有先获取并阅读之后，才能充分了解并理解论文所蕴含的信息与学术价值，并在成果中予以引证。

二者的强正相关关系，为其作为科研评价指标提供了有力论证。相较于引用，下载量具有实时以

及易获取等特点，弥补了引文指标存在时滞的缺陷。对于高被引论文，下载量可在早期发挥预测功能，对于低被引以及零被引论文则具有从“使用”角度进行评价的功能。围绕引用与下载数据，学者们提出了从全局与局部两种视角进行学术评价的思路。考虑到 Web of Science、Scopus 以及 CSSCI 等引文

数据库的存在,可以从全局视角提供论文、期刊、学者以及机构等多层次的评价数据。而 ScienceDirect 以及 CNKI 等数字图书馆、PLoS 等期刊网站以及机构库等提供的下载量数据都仅仅是学术文献集合的真子集,仅具有局部、微观的评价功能^[37]。

不过,也有学者指出,下载数据容易受到恶意自动下载程序的干扰而失真,所以在实践中应注意保障该数据的可靠性^[58]。此外,同一使用者可能多次下载论文导致下载数据膨胀,也是其存在的问题之一^[59]。

5.2 下载量与被引量相关关系的调节变量

本研究发现论文质量对二者相关关系存在较强的调节效应,论文质量越高,其下载后被引用的可能性也越大。引证时间窗、下载时间窗及下载数据源也存在一定的调节效应。

样本论文质量。本文用样本来源期刊的影响因子代表样本论文的质量,发现论文质量对下载量与被引量的相关关系具有显著调节效应。论著质量越高,下载后就越有可能被引用,相反,较低质量的论著,由于不能为研究者提供相应的方法借鉴或观点支撑等原因而被搁置。庞景安^[39]的研究发现影响因子前 100 名的期刊总下载量与总被引量的相关性较一般期刊更强,这与本研究的结论不谋而合。

下载数据源的调节效应也部分支持了样本论文质量的结论。ScienceDirect 以及 CNKI 数据库中的论文经过了同行评议,从而保证了论文质量;而搜索引擎以及开放存取平台的论文多未经同行评议,难以保证整体的高质量;也就是说,学术数据库中的论文质量总体高于后者,即从前者到后者相关关系逐次递减的内在原因更多的是样本论文质量的体现,当然也不能排除学术数据库易于访问、对期刊的覆盖更为全面等因素在起作用。

本研究结论的启示价值是:科研人员应严把论文质量关,选题新颖、方法严谨、逻辑缜密,使读者可以很快发现其学术价值,从而将潜在的、间接的“使用”,转换为实质的、直接的“引用”。目前,尚未见不同质量论文的下载量与被引量相关性的实证研究,可作为下一步的研究课题。

引用时间窗和下载时间窗。下载量信息具有实时性,而被引量则是论文发表一段时间之后被正式引证的产物,论文需要经过 2~3 年的同行评议直至发表,因而引用较下载滞后^[45]。可见,论文发表早期下载量增长快,随着时间不断增加,被引量的累

积效应更大,故引用时间窗和下载时间窗对两者的相关性具有调节效应。本研究表明,论文发表后,下载量与被引量的相关性逐渐增加,并在发表后第三年达到第一个高峰,随后其相关关系降低,但在发表五年后,两者的相关关系再次回升,该结论与方红玲^[8]和 Schlögl 等^[28]的研究一致。时间窗 ≥ 7 年之后,下载量与被引量的相关系数趋于稳定(0.7 左右),其原因可以从 JCR 的 IF2 以及 IF5 得到启示,即大部分论文已跃过半衰期而鲜有新的下载与引用所致^[12-14]。

下载数据源。学术数据库、搜索引擎或期刊网站等平台学科覆盖范围的不同;题录、文摘或者全文等服务方式的不同;PDF、html 以及 XML 等下载格式的不同影响了信息用户的访问与存取,从而导致不同平台被引量与下载量相关关系的差异。本文发现,学术数据库(如 ScienceDirect 和 CNKI)中论文的下载与被引具有较强的相关关系,而期刊网站和开放存取平台中两者的相关性较弱。为了引导使用数据从局部走向全局,论文下载数据统一记录平台的建立将是必要的,将大大增加其作为学术评价指标的可行性。本研究提示科研人员应努力提高论文在学术检索系统中的可见性,可达增大下载量及增加被引和学术影响力之效。目前,还未见比较不同数据源中下载数据特点与性质的实证研究,相关工作可据此展开。

本文没有发现样本来源学科和被引数据源对下载量与被引量相关关系的影响。Vaughan 等^[60]的研究虽发现不同学科论文间下载量与被引量的相关关系具有差异,但由于研究的样本选择范围大、涉及调节变量多,学科的调节效应可能由其他变量所引起。O'Leary^[45]通过研究发现,SSCI、Elsevier 以及 Google Scholar 搜索引擎的被引量高度相关($r=0.890-0.990$),提示被引数据源的差异对下载量与被引量相关关系的影响可以忽略。

元分析的质量受到原始研究质量和可获取性的限制^[61]。在本元分析中,超过半数的原始研究没有报告样本的文献类型,因而本研究难以探讨文献类型对下载量与被引量相关关系的调节效应。

参 考 文 献

(带*号为纳入研究的文献)

- [1] Borgman C L, Furner J. Scholarly communication and bibliometrics[J]. Annual Review of Information Science and Technology, 2002, 36(1): 2-72.
- [2] Davis J B. Problems in using the Social Sciences Citation Index to rank economics journals[J]. The American Economist,

- 1998, 42(2): 59-64.
- [3] Kurtz M J, Bollen J. Usage bibliometrics[J]. *Annual Review of Information Science and Technology*, 2010, 44(1): 1-64.
 - [4] Duy J, Vaughan L. Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination[J]. *The Journal of Academic Librarianship*, 2006, 32(5): 512-517.
 - [5] Wang X W, Fang Z H, Sun X L. Usage patterns of scholarly articles on Web of Science: a study on Web of Science usage count[J]. *Scientometrics*, 2016, 109(2): 917-926.
 - [6]* Li C L, Xu Y Q, Wu H, et al. Correlation and interaction visualization of altmetric indicators extracted from scholarly social network activities: Dimensions and structure[J]. *Journal of Medical Internet Research*, 2013, 15(11): e259.
 - [7] Hammook Z, Mišić J, Mišić V B. Student/supervisor collaboration and usage patterns of publications available on researchgate[C]// *Proceedings of 2016 IEEE Wireless Communications and Networking Conference*. IEEE, 2016: INSPEC Accession No. 16303660.
 - [8] 方红玲. 我国科技期刊论文被引量 and 下载量峰值年代——多学科比较研究[J]. *中国科技期刊研究*, 2011, 22(5): 708-710.
 - [9] 刘筱敏, 张建勇. 数字资源获取对科学研究的影响——电子期刊全文下载与引用分析[J]. *大学图书馆学报*, 2009, 27(1): 60-63.
 - [10] 万锦堃, 花平寰, 孙秀坤. 期刊论文被引用及其 Web 全文下载的文献计量分析[J]. *现代图书情报技术*, 2005(4): 58-62.
 - [11] Moed H F, Halevi G. On full text download and citation distributions in scientific-scholarly journals[J]. *Journal of the Association for Information Science and Technology*, 2016, 67(2): 412-431.
 - [12]* Moed H F. Statistical relationships between downloads and citations at the level of individual documents within a single journal[J]. *Journal of the American Society for Information Science and Technology*, 2005, 56(10): 1088-1097.
 - [13] Nicholas D, Huntington P, Dobrowolski T, et al. Revisiting 'obsolescence' and journal article 'decay' through usage data: an analysis of digital journal use by year of publication[J]. *Information Processing & Management*, 2005, 41(6): 1441-1461.
 - [14] Kurtz M J, Eichhorn G, Accomazzi A, et al. The bibliometric properties of article readership information[J]. *Journal of the American Society for Information Science and Technology*, 2005, 56(2): 111-128.
 - [15]* 丁佐奇, 郑晓南, 吴晓明. 科技论文被引频次与下载频次的相关性分析[J]. *中国科技期刊研究*, 2010, 21(4): 467-470.
 - [16]* Brody T, Harnad S, Carr L. Earlier Web usage statistics as predictors of later citation impact[J]. *Journal of the American Society for Information Science and Technology*, 2006, 57(8): 1060-1072.
 - [17]* Jahandideh S, Abdolmaleki P, Asadabadi E B. Prediction of future citations of a research paper from number of its internet downloads[J]. *Medical Hypotheses*, 2007, 69(2): 458-459.
 - [18]* Botting N, Dipper L, Hilari K. The effect of social media promotion on academic article uptake[J]. *Journal of the Association for Information Science and Technology*, 2017, 68(3): 795-800.
 - [19] 郭强, 赵瑾, 刘新新, 等. 利用期刊下载次数估计后期被引次数的研究[J]. *图书馆理论与实践*, 2010(11): 45-49.
 - [20] Pinkowitz L. Research dissemination and impact: Evidence from Web site downloads[J]. *The Journal of Finance*, 2002, 57(1): 485-499.
 - [21] Appell H J. Is the future of scientific journals electronic? Some considerations about downloads and citations[J]. *International Journal of Sports Medicine*, 2007, 28(11): 899-900.
 - [22]* Nieder C, Dalhaug A, Aandahl G. Correlation between article download and citation figures for highly accessed articles from five open access oncology journals[J]. *SpringerPlus*, 2013, 2(1): 261.
 - [23]* Subotic S, Mukherjee B. Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles[J]. *Journal of Information Science*, 2014, 40(1): 115-124.
 - [24]* 接雅俐, 郭立锦. 八种理工及农业类院校学报社科论文被引与下载情况分析[J]. *合肥工业大学学报(社会科学版)*, 2011, 25(4): 142-146.
 - [25] Lippi G, Favaloro E J. Article downloads and citations: Is there any relationship?[J]. *Clinica Chimica Acta*, 2013, 415: 195.
 - [26]* Naudé F. Comparing downloads, Mendeley readership and Google Scholar citations as indicators of article performance [J]. *The Electronic Journal of Information Systems in Developing Countries*, 2017, 78(1): 1-25.
 - [27] Glänzel W, Gorraiz J. Usage metrics versus altmetrics: confusing terminology?[J]. *Scientometrics*, 2015, 102(3): 2161-2164.
 - [28]* Schlögl C, Gorraiz J, Gumpenberger C, et al. Comparison of downloads, citations and readership data for two information systems journals[J]. *Scientometrics*, 2014, 101(2): 1113-1128.
 - [29]* Schloegl C, Gorraiz J. Comparison of citation and usage indicators: the case of oncology journals[J]. *Scientometrics*, 2010, 82(3): 567-580.
 - [30]* Wan J K, Hua P H, Rousseau R, et al. The journal download immediacy index (DII): experiences using a Chinese full-text database[J]. *Scientometrics*, 2010, 82(3): 555-566.
 - [31]* Shuai X, Pepe A, Bollen J. How the scientific community reacts to newly submitted preprints: article downloads, Twitter mentions, and citations[J]. *PLoS ONE*, 2012, 7(11): e47523.
 - [32] Hajjem C, Harnad S, Gingras Y. Ten-year cross-disciplinary

- comparison of the growth of open access and how it increases research citation impact[J]. *Computer Science*, 2006, 2013(4): 39-47.
- [33] Wang X W, Liu C, Mao W L, et al. The open access advantage considering citation, article usage and social media attention[J]. *Scientometrics*, 2015, 103(2): 555-564.
- [34]* Jamali H R, Nikzad M. Article title type and its relation with the number of downloads and citations[J]. *Scientometrics*, 2011, 88(2): 653-661.
- [35] Guerrero-Bote V P, Moya-Anegón F. Relationship between downloads and citations at journal and paper levels, and the influence of language[J]. *Scientometrics*, 2014, 101(2): 1043-1065.
- [36] 牛昱昕, 宗乾进, 袁勤俭. 开放存取论文下载与引用情况计量研究[J]. *中国图书馆学报*, 2012, 38(4): 119-127.
- [37] Bollen J, Van de Sompel H. Usage impact factor: The effects of sample characteristics on usage-based impact metrics[J]. *Journal of the American Society for Information Science and Technology*, 2008, 59(1): 136-149.
- [38] Gorraiz J, Gumpenberger C, Schlögl C. Usage versus citation behaviours in four subject areas[J]. *Scientometrics*, 2014, 101(2): 1077-1095.
- [39] 庞景安. 中文科技期刊下载计量指标与引用计量指标的比较研究[J]. *情报理论与实践*, 2006, 29(1): 44-48.
- [40] Kurtz M J, Henneken E A. Measuring metrics—a 40-year longitudinal cross validation of citations, downloads, and peer review in astrophysics[J]. *Journal of the Association for Information Science and Technology*, 2017, 68(3): 695-708.
- [41] Boukacem-Zeghmouri C, Bador P, Lafouge T, et al. Relationships between consumption, publication and impact in French universities in a value perspective: a bibliometric analysis[J]. *Scientometrics*, 2016, 106(1): 263-280.
- [42]* Liu X L, Fang H L, Wang M Y. Correlation between download and citation and download-citation deviation phenomenon for some papers in Chinese medical journals[J]. *Serials Review*, 2011, 37(3): 157-161.
- [43]* Schloegl C, Gorraiz J. Global usage versus global citation metrics: The case of pharmacology journals[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(1): 161-170.
- [44] Ketcham C M. The proper use of citation data in journal management[J]. *Archivum Immunologiae et Therapiae Experimentalis*, 2009, 56(6): 357-362.
- [45]* O'Leary D E. The relationship between citations and number of downloads in *Decision Support Systems*[J]. *Decision Support Systems*, 2008, 45(4): 972-980.
- [46] Hunter J E, Schmidt F L. *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.)[M]. Thousand Oaks: SAGE Publications, 2004.
- [47] Peterson R A, Brown S P. On the use of beta coefficients in meta-analysis[J]. *Journal of Applied Psychology*, 2005, 90(1): 175-181.
- [48] Rosenthal R. *Meta-analytic procedures for social research*[M]. Thousand Oaks: SAGE Publications, 1991.
- [49] Bornmann L. Alternative metrics in scientometrics: a meta-analysis of research into three altmetrics[J]. *Scientometrics*, 2015, 103(3): 1123-1144.
- [50] Hunter J E, Schmidt F L. Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge[J]. *International Journal of Selection and Assessment*, 2000, 8(4): 275-292.
- [51] Lipsey M W, Wilson D B. *Practical meta-analysis*[M]. Thousand Oaks: SAGE Publications, 2000.
- [52] Cohen J. *Statistical power analysis for the behavioral sciences*[M]. Psychology Press, 1988.
- [53]* 杜秀杰, 赵大良, 葛赵青, 等. 学术论文的下载频率与被引频率的相关性分析[J]. *编辑学报*, 2009, 21(6): 551-553.
- [54] Higgins J P T, Thompson S G, Deeks J J, et al. Measuring inconsistency in meta-analyses[J]. *BMJ*, 2003, 327(7414): 557-560.
- [55] Duval S, Tweedie R. Trim and Fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis[J]. *Biometrics*, 2000, 56(2): 455-463.
- [56] 陆伟, 钱坤, 唐祥彬. 文献下载频次与被引频次的相关性研究——以图书情报领域为例[J]. *情报科学*, 2016, 34(1): 3-8.
- [57] 曹艺, 王曰芬, 丁洁. 面向学术影响力评价的科技文献引用与下载的相关性研究[J]. *图书情报工作*, 2012, 56(8): 56-64.
- [58] Neylon C, Wu S. Article-level metrics and the evolution of scientific impact[J]. *PLoS Biology*, 2009, 7(11): e1000242.
- [59] Taraborelli D. Soft peer review: Social software and distributed scientific evaluation[C]// *Proceedings of the 8th International Conference on the Design of Cooperative Systems*, Carry-Le-Rouet, 2008.
- [60]* Vaughan L, Tang J, Yang R. Investigating disciplinary differences in the relationships between citations and downloads[J]. *Scientometrics*, 2017, 111(3): 1533-1545.
- [61] Ellis G, Whitehead M A, Robinson D, et al. Comprehensive geriatric assessment for older adults admitted to hospital: meta-analysis of randomised controlled trials[J]. *BMJ*, 2011, 343: d6553.
- 以下为纳入元分析研究但未在正文中出现的文献:
- * O'Leary D E. Downloads and citations in *Intelligent Systems in Accounting, Finance and Management*[J]. *Intelligent Systems in Accounting, Finance and Management*, 2009, 16(1-2): 21-31.
- * Chu H, Krichel T. Downloads vs. Citations: relationships,

contributing factors and beyond[J]. International Society for Scientometrics & Informetrics, 2007.

* O’Leary D E. On the relationship between citations and appearances on “top 25” download lists in the International Journal of Accounting Information Systems[J]. International Journal of Accounting Information Systems, 2008, 9(1): 61-75.

* Avello Martínez R, Anderson T. Are the most highly cited articles the ones that are the most downloaded? A bibliometric study of IRRODL[J]. The International Review of Research in Open and Distributed Learning, 2015, 16(3): 18-40.

* Thelwall M, Kousha K. Do journal data sharing mandates

work? Life sciences evidence from Dryad[J]. Aslib Journal of Information Management, 2017, 69(1): 36-45.

* 王雅祺. 下载数与被引量关系及其影响因素分析[D]. 天津: 天津工业大学, 2016.

* 邢星. 论文下载量与其被引量间关系的初步研究[D]. 沈阳: 中国医科大学, 2013.

* 孔秋丽. 期刊影响因子、论文被引次数和下载次数数的关系研究[D]. 桂林: 广西师范大学, 2015.

* 马云彤. 高被引论文被引频次与下载频次的相关性分析——基于中国知网以 2006-2010 年出版专题期刊论文为分析对象[J]. 西安文理学院学报: 自然科学版, 2012, 15(4): 122-124.

附录

附录 样本编码结果

| 作者及发表年份 | 样本发表年份 | 样本量 | 样本来源学科 | 样本来源期刊 | 样本质量 | 被引时间窗 | 下载时间窗 | 被引数据源 | 下载数据源 | 相关系数 |
|-----------------|-----------|-------|-----------|--|------|-------|-------|--------|----------------|-------|
| Jamali2011 | 2007 | 2172 | 生命科学与生物医学 | PLoS 的 6 本期刊 | 1 | 3 | 3 | 期刊网站 | 期刊网站 | 0.550 |
| Schloegl2010(1) | 2004 | 145 | 生命科学与生物医学 | Cancer Cell | 1 | 2 | 2 | Scopus | Science-Direct | 0.510 |
| Schloegl2010(2) | 2001 | 328 | 生命科学与生物医学 | Cancer Letters | 3 | 5 | 5 | Scopus | Science-Direct | 0.630 |
| Schloegl2010(3) | 2002 | 294 | 生命科学与生物医学 | Gynecologic Oncology | 2 | 4 | 4 | Scopus | Science-Direct | 0.610 |
| Schloegl2014(1) | 2002-2011 | 150 | 应用科学 | Journal of Strategic Information Systems | 2 | — | — | Scopus | Science-Direct | 0.770 |
| Schloegl2014(2) | 2002-2011 | 528 | 应用科学 | Information and Management | 1 | — | — | Scopus | Science-Direct | 0.760 |
| Liu2011 | 2005 | 1622 | 生命科学与生物医学 | 5 本中文期刊 | 4 | 4 | 4 | CNKI | CNKI | 0.491 |
| Liu2013 | — | 33128 | 生命科学与生物医学 | — | — | — | — | Scopus | 期刊网站 | 0.444 |
| Xin2012 | 2010-2011 | 70 | 自然科学 | — | — | 1 | 1 | 搜索引擎 | 开放存取 | 0.387 |
| Brody2006 | — | 14200 | 自然科学 | — | — | 2 | 2 | 开放存取 | 开放存取 | 0.440 |
| Moed2005 | 2001 | 1190 | 自然科学 | Tetrahedron Letters | 2 | 2 | 2 | WoS | Science-Direct | 0.220 |
| Botting2017 | 2005-2013 | 30 | 社会科学 | — | — | — | — | Scopus | 开放存取 | 0.450 |
| Wan2010 | 2004 | 460 | 生命科学与生物医学 | — | — | 2 | 1 | CNKI | CNKI | 0.635 |
| Schloegl2011(1) | 2001 | 111 | 生命科学与生物医学 | Drug Discovery Today | 1 | 5 | 5 | WoS | Science-Direct | 0.770 |
| Schloegl2011(2) | 2002 | 105 | 生命科学与生物医学 | Drug Discovery Today | 1 | 4 | 4 | WoS | Science-Direct | 0.730 |
| Schloegl2011(3) | 2003 | 75 | 生命科学与生物医学 | Drug Discovery Today | 1 | 3 | 3 | WoS | Science-Direct | 0.650 |
| Schloegl2011(4) | 2004 | 72 | 生命科学与生物医学 | Drug Discovery Today | 1 | 2 | 2 | WoS | Science-Direct | 0.560 |

续表

| 作者及发表年份 | 样本发表年份 | 样本量 | 样本来源学科 | 样本来源期刊 | 样本质量 | 被引时间窗 | 下载时间窗 | 被引数据来源 | 下载数据来源 | 相关系数 |
|------------------|-----------|------|-----------|--|------|-------|-------|--------|----------------|-------|
| Schloegl2011(5) | 2005 | 155 | 生命科学与生物医学 | Drug Discovery Today | 1 | 1 | 1 | WoS | Science-Direct | 0.680 |
| Nieder2013 | 2001-2012 | 250 | 生命科学与生物医学 | BMC 的 5 本期刊 | 2 | — | — | Scopus | 期刊网站 | 0.230 |
| O'Leary2009 | 1996-2002 | 48 | 社会科学 | Intelligent Systems in Accounting, Finance and Management(ISAFM) | 4 | — | — | 搜索引擎 | 期刊网站 | 0.437 |
| Chu2007(1) | — | 200 | 社会科学 | — | — | — | — | 搜索引擎 | 开放存取 | 0.610 |
| Chu2007(2) | — | 200 | 社会科学 | — | — | — | — | WoS | 开放存取 | 0.540 |
| O'Leary 2008a(1) | 2002-2004 | 25 | 应用科学 | Decision Support Systems | 2 | — | — | WoS | Science-Direct | 0.839 |
| O'Leary 2008a(2) | 2002-2004 | 25 | 应用科学 | Decision Support Systems | 2 | — | — | Scopus | Science-Direct | 0.820 |
| O'Leary 2008a(3) | 2002-2004 | 25 | 应用科学 | Decision Support Systems | 2 | — | — | 搜索引擎 | Science-Direct | 0.784 |
| O'Leary 2008b(1) | 2002-2006 | 25 | 社会科学 | International Journal of Accounting Information Systems | 3 | — | — | WoS | Science-Direct | 0.358 |
| O'Leary 2008b(2) | 2002-2006 | 25 | 社会科学 | International Journal of Accounting Information Systems | 3 | — | — | Scopus | Science-Direct | 0.462 |
| O'Leary 2008b(3) | 2002-2006 | 25 | 社会科学 | International Journal of Accounting Information Systems | 3 | — | — | 搜索引擎 | Science-Direct | 0.550 |
| Subotic2013 | — | 258 | 社会科学 | — | — | 2 | — | Scopus | Science-Direct | 0.220 |
| Avello2015 | 2008-2013 | 100 | 社会科学 | The International Review of Research in Open and Distance Learning | 3 | — | — | 搜索引擎 | 开放存取 | 0.621 |
| Jahandideh2007 | 2006 | 50 | 自然科学 | Journal of Chromatography A | 2 | — | — | Scopus | Science-Direct | 0.304 |
| Naude2017 | 2000-2013 | 378 | 应用科学 | EJISDC | 4 | — | — | 搜索引擎 | 期刊网站 | 0.660 |
| Thelwall2016(1) | 2012 | 169 | 生命科学与生物医学 | Evolution | 1 | 4 | 4 | Scopus | 开放存取 | 0.149 |
| Thelwall2016(2) | 2012 | 63 | 生命科学与生物医学 | Heredity | 1 | 4 | 4 | Scopus | 开放存取 | 0.181 |
| Vaughan2017(1) | 2011 | 1500 | 自然科学 | 30 种中文期刊 | — | 5 | 5 | CNKI | CNKI | 0.700 |
| Vaughan2017(2) | 2011 | 1500 | 应用科学 | 30 种中文期刊 | — | 5 | 5 | CNKI | CNKI | 0.720 |
| Vaughan2017(3) | 2011 | 1500 | 生命科学与生物医学 | 30 种中文期刊 | — | 5 | 5 | CNKI | CNKI | 0.630 |
| Vaughan2017(4) | 2011 | 1500 | 社会科学 | 30 种中文期刊 | — | 5 | 5 | CNKI | CNKI | 0.830 |
| Vaughan2017(5) | 2011 | 1500 | 艺术与人文科学 | 30 种中文期刊 | — | 5 | 5 | CNKI | CNKI | 0.750 |
| 王雅琪 2016 | 2000-2014 | 4512 | 应用科学 | — | — | — | — | CNKI | CNKI | 0.725 |
| 杜秀杰 2009(1) | 1994 卷 1 | 18 | 自然科学 | 西安交通大学学报 | 4 | — | — | CNKI | CNKI | 0.948 |
| 杜秀杰 2009(2) | 2006 卷 1 | 28 | 自然科学 | 西安交通大学学报 | 4 | — | — | CNKI | CNKI | 0.645 |
| 杜秀杰 2009(3) | 2006 卷 2 | 31 | 自然科学 | 西安交通大学学报 | 4 | — | — | CNKI | CNKI | 0.533 |
| 杜秀杰 2009(4) | 2006 卷 3 | 29 | 自然科学 | 西安交通大学学报 | 4 | — | — | CNKI | CNKI | 0.398 |
| 杜秀杰 2009(5) | 2007 卷 1 | 30 | 自然科学 | 西安交通大学学报 | 4 | — | — | CNKI | CNKI | 0.008 |

续表

| 作者及发表年份 | 样本发表年份 | 样本量 | 样本来源学科 | 样本来源期刊 | 样本质量 | 被引时间窗 | 下载时间窗 | 被引数据源 | 下载数据源 | 相关系数 |
|-------------|-----------|------|---------------|---------------------|------|-------|-------|-------|-------|--------|
| 邢星 2013(1) | 1995 | 1721 | 应用科学 | — | — | 17 | 17 | CNKI | CNKI | 0.727 |
| 邢星 2013(2) | 1995 | 249 | 生命科学与 生物医学 | — | — | 17 | 17 | CNKI | CNKI | 0.725 |
| 邢星 2013(3) | 1995 | 917 | 自然科学 | — | — | 17 | 17 | CNKI | CNKI | 0.611 |
| 邢星 2013(4) | 1995 | 4226 | 生命科学与 生物医学 | — | — | 17 | 17 | CNKI | CNKI | 0.548 |
| 邢星 2013(5) | 2000 | 3006 | 应用科学 | — | — | 12 | 12 | CNKI | CNKI | 0.807 |
| 邢星 2013(6) | 2000 | 496 | 生命科学与 生物医学 | — | — | 12 | 12 | CNKI | CNKI | 0.687 |
| 邢星 2013(7) | 2000 | 1357 | 自然科学 | — | — | 12 | 12 | CNKI | CNKI | 0.751 |
| 邢星 2013(8) | 2000 | 5065 | 生命科学与 生物医学 | — | — | 12 | 12 | CNKI | CNKI | 0.656 |
| 邢星 2013(9) | 2005 | 4578 | 应用科学 | — | — | 7 | 7 | CNKI | CNKI | 0.807 |
| 邢星 2013(10) | 2005 | 765 | 生命科学与 生物医学 | — | — | 7 | 7 | CNKI | CNKI | 0.643 |
| 邢星 2013(11) | 2005 | 1719 | 自然科学 | — | — | 7 | 7 | CNKI | CNKI | 0.740 |
| 邢星 2013(12) | 2005 | 6046 | 生命科学与 生物医学 | — | — | 7 | 7 | CNKI | CNKI | 0.632 |
| 邢星 2013(13) | 2010 | 7564 | 应用科学 | — | — | 2 | 2 | CNKI | CNKI | 0.591 |
| 邢星 2013(14) | 2010 | 799 | 生命科学与 生物医学 | — | — | 2 | 2 | CNKI | CNKI | 0.539 |
| 邢星 2013(15) | 2010 | 2467 | 自然科学 | — | — | 2 | 2 | CNKI | CNKI | 0.684 |
| 邢星 2013(16) | 2010 | 5027 | 生命科学与 生物医学 | — | — | 2 | 2 | CNKI | CNKI | 0.439 |
| 孔秋丽 2015(1) | 2010 | 101 | 自然科学 | — | — | 4 | 4 | CNKI | CNKI | 0.882 |
| 孔秋丽 2015(2) | 2011 | 58 | 自然科学 | — | — | 3 | 3 | CNKI | CNKI | 0.791 |
| 孔秋丽 2015(3) | 2012 | 152 | 自然科学 | — | — | 2 | 2 | CNKI | CNKI | 0.599 |
| 马云彤 2012(1) | 2006-2010 | 50 | 社会科学 | — | — | — | — | CNKI | CNKI | 0.341 |
| 马云彤 2012(2) | 2006-2010 | 50 | 社会科学 | — | — | — | — | CNKI | CNKI | 0.500 |
| 马云彤 2012(3) | 2006-2010 | 50 | 社会科学 | — | — | — | — | CNKI | CNKI | 0.335 |
| 丁佐奇 2010(1) | 2003-2008 | 20 | 生命科学与 生物医学 | 《中国天然药物》 | 3 | — | — | CNKI | CNKI | 0.331 |
| 丁佐奇 2010(2) | 2003-2008 | 20 | 生命科学与 生物医学 | 《中国药科大学学报》 | 4 | — | — | CNKI | CNKI | 0.012 |
| 接雅俐 2011(1) | 2005-2009 | 10 | 社会科学 | 南京农业大学学报 (社会科学版) | — | — | — | CNKI | CNKI | 0.390 |
| 接雅俐 2011(2) | 2005-2009 | 10 | 社会科学 | 南京农业大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | -0.359 |
| 接雅俐 2011(3) | 2005-2009 | 10 | 社会科学 | 中国农业大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.427 |
| 接雅俐 2011(4) | 2005-2009 | 10 | 社会科学 | 中国农业大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | -0.231 |
| 接雅俐 2011(5) | 2005-2009 | 10 | 社会科学 | 中国地质大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.191 |

续表

| 作者及发表年份 | 样本发表年份 | 样本量 | 样本来源学科 | 样本来源期刊 | 样本质量 | 被引时间窗 | 下载时间窗 | 被引数 据源 | 下载数 据源 | 相关 系数 |
|--------------|-----------|-----|--------|---------------------|------|-------|-------|-----------|-----------|----------|
| 接雅俐 2011(6) | 2005-2009 | 10 | 社会科学 | 中国地质大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.644 |
| 接雅俐 2011(7) | 2005-2009 | 10 | 社会科学 | 大连理工大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | -0.079 |
| 接雅俐 2011(8) | 2005-2009 | 10 | 社会科学 | 大连理工大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | -0.127 |
| 接雅俐 2011(9) | 2005-2009 | 10 | 社会科学 | 华中科技大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.579 |
| 接雅俐 2011(10) | 2005-2009 | 10 | 社会科学 | 华中科技大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.382 |
| 接雅俐 2011(11) | 2005-2009 | 10 | 社会科学 | 华南农业大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.536 |
| 接雅俐 2011(12) | 2005-2009 | 10 | 社会科学 | 华南农业大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.470 |
| 接雅俐 2011(13) | 2005-2009 | 10 | 社会科学 | 北京理工大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.463 |
| 接雅俐 2011(14) | 2005-2009 | 10 | 社会科学 | 北京理工大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.494 |
| 接雅俐 2011(15) | 2005-2009 | 10 | 社会科学 | 北京交通大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.317 |
| 接雅俐 2011(16) | 2005-2009 | 10 | 社会科学 | 北京交通大学学报 (社会科学版) | 4 | — | — | CNKI | CNKI | 0.406 |

(责任编辑 魏瑞斌)