

美捷登精彩点评 谷歌学术 H5 指数及其与汤森路透影响因子对比分析

Mike Wang

每年汤森路透 (Thomson Reuters) 的期刊引用报告 (Journal Citation Report, JCR) 在中国都引起了高度的关注。原因是在中国科研评价中, 科研论文是否发表在被科学引文索引 (Science Citation Index, SCI) 或科学引文索引扩展版 (Science Citation Index Expanded, SCI-E) 收录的杂志上, 以及该杂志的影响因子的高低都起着举足轻重的作用, 甚至成为某些部门和单位唯一的科研评价指标。近年来, 依赖汤森路透 JCR 的科研评价系统受到了质疑、争议、批评和诟病, 因而出现了寻求新的科研评价系统的呼声。

谷歌学术 (Google Scholar) 在 2012 年推出了一个杂志评价系统, 即谷歌学术计量 (Google Scholar Metrics), 用来评价各个领域杂志的影响力。该系统主要包括 H 指数 (h-index 或 Hirsch index), H 核心 (h-core),

H 中值 (h-median), H5 指数 (h5-index), H5 核心 (h5-core) 和 H5 中值 (h5-median, 图 1)。

H 指数于 2005 年由美国加利福尼亚大学圣地亚哥分校物理学家乔治·希尔施 (Jorge Hirsch) 首先提出, 原本用来评价某一个学者的影响力, 计算基于论文总数及各论文被引用的次数, 即某学者如果在其所有学术论文中至少有 N 篇论文分别被引用了至少 N 次, 那么该学者的 H 指数就是 N。例如, 王教授共发表 200 篇论文, 其中有至少 50 篇被引用至少 50 次, 那么他的 H 指数就是 50。

在谷歌学术计量系统, 杂志的 H 指数指在所有发表的论文中有至少 h 篇论文分别被引用了至少 h 次, 那么这份杂志的 H 指数就是 h。例如, 某杂志共发表 5 篇论文, 分别被引用 17, 9, 6, 3 和 2 次, 那么这份杂志

的 H 指数就是 3; H 核心指该杂志最高被引用的 h 篇论文, 如上述杂志的 H 核心包括被引用 17, 9 和 6 次的三篇论文; H 中值指 H 核心中位数论文的引用次数, 如上述杂志的 H 中值为 9 (图 1)。

相应地, H5 指数, H5 核心和 H5 中值的计算基于收录在谷歌学术系统中的杂志最近 5 年的论文数量及各论文被引用的次数。例如, 如果某杂志在过去 5 年内所发表的论文中至少有 h 篇论文分别被引用了至少 h 次, 那么这份杂志的 H5 指数就是 h。H5 核心和 H5 中值依上述方法计算。谷歌学术计量报告虽然公布各杂志的 H5 指数, H5 核心和 H5 中值, 但最重要的还是 H5 指数。值得一提的是, 只有那些在 5 年内发表了 100 篇以上论文并且至少有一个引用的杂志才会被收录到谷歌学术中去。

出版物	h5 指数	h5 中位数
1. The New England Journal of Medicine	328	520
2. The Lancet	258	415
3. Cell	216	330
4. Proceedings of the National Academy of Sciences	216	280
5. Journal of Clinical Oncology	202	296
6. JAMA: The Journal of the American Medical Association	184	277
7. Circulation	182	278
8. Nature Genetics	182	268
9. Journal of the American College of Cardiology	163	244
10. PLoS ONE	161	210
11. Nature Medicine	153	228
12. Blood	152	202
13. British Medical Journal	143	191
14. The Cochrane Database of Systematic Reviews	140	207
15. Neuron	139	198
16. Gastroenterology	138	180
17. Cancer Research	138	179
18. The Lancet Oncology	137	228
19. The Journal of Clinical Investigation	133	180
20. Nature Reviews Cancer	131	250

图 1. Google 学术官网显示的健康与医疗科学刊物的 Top20

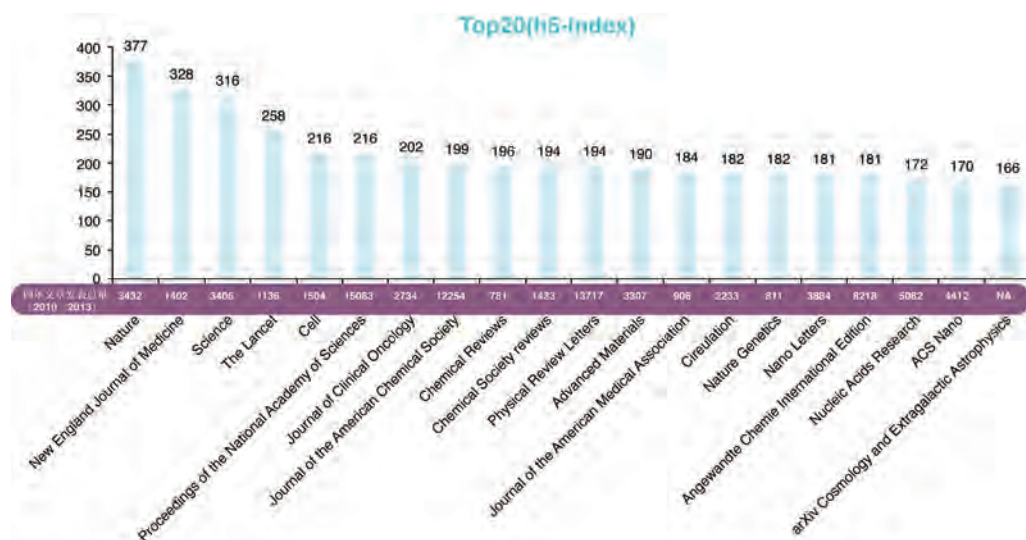


图 2. Top20 杂志

一、谷歌学术计量报告简介

到目前为止，谷歌已经连续四年发布了杂志学术计量报告。今年（2015 年）所发布的报告共包括 7211 份杂志，由于其中有 1761 份杂志在不同领域排名中重复出现，因此该报告中实际上只收录了 5450 份杂志。报告中给出了 9 种不同语言（英语、中文、葡萄牙语、德语、西班牙语、法语、意大利语、日语、荷兰语）杂志的前 100 名的信息。另外，该报告对英文类杂志的统计更详细，包括八大领域（商务、经济与管理、化学与材料科学、工程与计算机科学、健康与医学科学、人文、文学与艺术、生命科学和地球科学、物理与数学、社会科学）及 261 个学科杂志的详细分类排名。每学科只有 top20 杂志信息，所以没有其余中低档次杂志的排名信息。

谷歌学术指标系统中还提供了搜索功能，输入杂志名称关键词就会给出包含该关键词前 20 名杂志的排名信息。但是，与汤森路透的 JCR 数据库不同的是：JCR 数据库有往年的报告信息而谷歌学术计量报告没有往年报告信息，所以这两者之间不能进行横向比较。

二、英文杂志 Top20 分析

Nature（自然）杂志以 H5 指数 377（H5 中值为 529），排名第一（图 2），也就是在 2010 年到 2014 年这 5 年期间，*Nature* 发表的所有论文中引用数不低于 377 次的共有至少 377 篇论文；而在今年汤森路透的 JCR 报告中，*Nature* 以 41.456 的影响因子（Impact factor）排名第 7。

New England Journal of Medicine（新英格兰医学杂志）以 H5 指数 328（H5 中值为 520），排在第二。这个排名和汤森路透 JCR 报告的排名一致。

Science（科学）杂志，H5 指数为 316（H5 中值为 446），排名是第三位；而该杂志在汤森路透 JCR 报告中排在第 16 位。

著名 *Lancet*（柳叶刀）杂志以其 258（H5 中值为 415）的 H5 指数排在第四位，其排名也和汤森路透 JCR 报告中排名一致。

Cell（细胞）杂志和 *Proceedings of the National Academy of Sciences*（PNAS，美国国家科学院刊）杂志以 216 的 H5 指数并列排在第五位；而它们的 H5 中值数据不同，分别为 330 和 280。另外，这两个杂志在汤

森路透 JCR 报告中影响因子分别排在第 20 和 186 位。

其他进入 Top20 的杂志 H5 指数见图 1。其中排在第十八位的 *Nucleic Acids Research* 杂志在汤森路透 JCR 报告中排在第 213 位。有趣的是，排在汤森路透 JCR 报告中第一名的超牛杂志 *CA: A Cancer Journal for Clinicians*，其影响因子为 115.84，而其 H5 指数为 57，没有排进前 100 名。

三、H5 核心论文发表时间分析

H5 指数是对某杂志过去 5 年期间所发表的论文引用数进行的评价，例如，2015 年对 2010 ~ 2014 期间所发表对论文引用数进行的评价。笔者根据汤森路透 JCR 报告中的影响因子，选择了五份高中低档次发杂志，对 H5 核心论文发表时间进行力分析。这五份杂志分别为 *Nature*，*CA: A Cancer Journal for Clinicians*，*Cancer Research*，*PLoS ONE* 和 *Cutis*，他们的 H5 指数分别为 377、57、138、161、16（图 3）。

Nature: 对 *Nature* 杂志的这 377 篇论文发表时间分析发现，其中 2010 年有 163 篇（43.24%），2011 年有 128 篇（33.95%），2012 年有 76 篇（20.16%），而 2013 和

2014 年各有 9 篇（2.39%）和 1 篇（0.27%）（图 3A）。也就是说，该杂志 H5 指数的 377 篇论文主要是来自于 2010 ~ 2012 这三年（97.34%）；2013 和 2014 年的论文比例仅为 2.66%。

CA: A Cancer Journal for Clinicians:

对影响因子排名第一的综述性牛刊 *CA: A Cancer Journal for Clinicians* 的 57 篇论文发表时间分析发现，其中 2010 年有 16 篇（28.07%），2011 年有 17 篇（29.82%），2012 年有 12 篇（21.05%），2013 年有 7 篇（12.28%），以及 2014 年有 5 篇（8.77%）（图 3B），综述性牛刊的 57 篇论文在 2010 ~ 2012 三年的占 78.95%，而后两年 2013 和 2014 年的论文所占比例“高达”20% 之多（相较其他四份杂志而言）。

Cancer Research:

对 *Cancer Research* 杂志的 138 篇论文分析发现，2010 年有 77 篇（55.80%），2011 年有 41 篇（29.71%），2012 年有 14 篇（10.14%），而 2013 和 2014 年各有 5 篇（3.62%）和 1 篇（0.72%）（图 3C）。从中看出，前三年论文数占 95.65%，2013 和 2014 两年仅为 4.35%。

PLoS ONE:

PLoS ONE 杂志的 161 篇论文，其中 2010 年有 91 篇（56.52%），

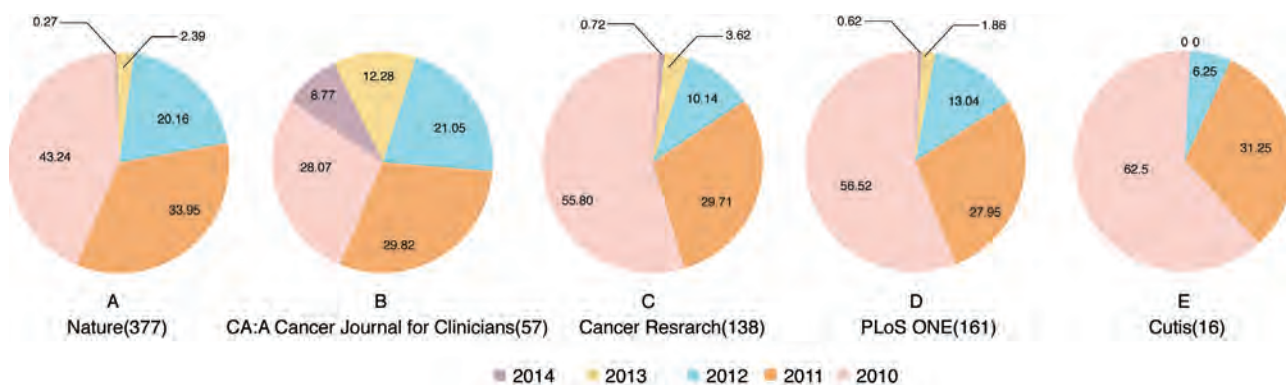


图 3. *Nature*，*CA: A Cancer Journal for Clinicians*，*Cancer Research*，*PLoS ONE* 和 *Cutis* 杂志 H5 核心论文发表时间分析

个别杂志的影响因子很高，而 H5 指数相对偏低；有的杂志则相反，H5 指数很高但影响因子却相对偏低；也有些杂志虽然影响因子不同，但是有相同的或相近的 H5 指数，比如 *Nano Research* 和 *CA: A Cancer Journal for Clinicians* 的 H5 指数都为 57，而它们的影响因子分别为 7.01 和 115.84，相差甚远。这是由于在过去的五年内，*Nano Research* 发表的论文总量为 500 多篇，是 *CA: A Cancer Journal for Clinicians* 杂志论文发表总量的 5 倍，也难怪如此低影响因子的杂志会和 *CA: A Cancer Journal for Clinicians* 有相同的 H5 指数，即 *Nano Research* 的 500 篇论文中有 57 篇论文的单篇引用数超过 57 次，而 *CA: A Cancer Journal for Clinicians* 的 100 篇论文中也有 57 篇论文的单篇引用数超过 57 次；就算 *CA: A Cancer Journal for Clinicians* 的全部（100 篇左右）论文的单篇引用数都超过 100，那它的 H5 指数也就才 100，挤进 H5 指数排名的 Top100 都有困难。

同样，*PLoS ONE* 的 H5 指数为 161，即在该杂志近五年内发表的 10.6 万多篇（106936）论文中，有 161 篇（占总发表量的 0.15%）的单篇引用数超过 161 次。H5 指数似乎更强调“大网捞鱼”，论文发表数量多的杂志很占优势。笔者从汤森路透官网收集了 Top20 杂志的 2010 ~ 2013 四年的论文发表总量，如图 2 数据显示：Top20 杂志多数既有较高的“论文发表总量”，同时也有较高的“单篇论文引用数”；这两者犹如天平的两端，像 *CA: A Cancer Journal for Clinicians* 这样的牛杂志，由于总论文发表量低，导致其 H5 指数偏低。

也有少数杂志是影响因子很高，H5 指数也很高，比如新英格兰医学杂志（图 4 红框内最上面的杂志），前提是该杂志论文发

表数量要足够多。谷歌学术 H5 指数前 100 名杂志的汤森路透 JCR 排名和影响因子见附表。

五、谷歌学术 H5 指数的优缺点

优点

1. 汤森路透集团 JCR 报告需要购买账号查询，而谷歌学术 H5 指数查询是免费的。
2. H5 指数对 9 种不同语言的出版物前一百名进行评估分析，因此有广泛代表性。
3. H5 指数不受有超高引用单篇论文的影响，相对比较客观。
4. 谷歌学术数据库收入范围非常广泛，不仅包括学术期刊，而且还收录书籍、会议、以及各种预印本。

缺点

1. 主要评估杂志近五年内，头三年所发表论文的引用情况。
2. 偏重年发表论文量大的杂志。
3. 由于不能查看谷歌的 H5 指数往年数据，所以不能进行年度横向对比（除非有心人每年都存下当年数据）。
4. 谷歌学术指标数据库只提供了英文杂志各领域前 20 名杂志信息，其余中低档次杂志信息未知。
5. 谷歌没有提供对各领域及学科分类的标准，也没有提供每个杂志所隶属的学科信息。

综上所述，谷歌学术指标具有一定的代表性和指导意义，但由于谷歌的学术指标才发布了四年，有些问题仍有待完善，比如杂志数据库收录的各项指标等，一旦修改就会影响全部数据。■

Google