● 张逸勤[1,2]，杨　杰[1,2]，宋欣雨[1,2]，邓三鸿[1,2]，赵　星[3,4]
（1. 南京大学信息管理学院，江苏　南京　210023；2. 数据工程与知识服务省高校重点实验室（南京大学），江苏　南京　210023；3. 复旦大学大数据研究院，上海　200433；4. 复旦大学国家智能评价与治理实验基地，上海　200433）

# 基于代表作的科学家学术影响力评价模式研究*
## ——以诺贝尔奖得主为例

**摘　要**：［目的/意义］学术评价在人才遴选、项目评审和科技政策制定等方面发挥着关键作用。随着“破五唯”改革深入，代表作评价制度作为一种新型评价模式日益受到关注，但目前仍缺乏系统的定量分析支撑。［方法/过程］提出以“代表作篇均被引量”作为科学家学术影响力的定量表征。基于微软学术数据库(MAG)，对物理学、化学和生理学或医学领域的301位诺贝尔奖得主及2518位对照组科学家进行实证分析。通过相关性分析、均值差异比较、识别率、平均排名及逻辑回归等多维度验证该评价模式的有效性。［结果/结论］①基于代表作的评价模式相较于传统的h指数等计量指标具有更高的区分效度；②该指标与诺奖得主评价的专家判断具有较高一致性；③该评价模式在不同学科领域均展现出良好的适用性。文章为科学家学术影响力的量化评价提供了新的理论框架和实证依据，对推进学术评价体系改革具有重要的理论价值和实践指导意义。［局限］主要针对自然科学领域成熟科学家群体，对人文社科领域及青年学者的适用性有待进一步验证。

**关键词**：代表作；诺贝尔奖；引文分析；学术评价

**DOI**：10. 16353/j. cnki. 1000-7490. 2025. 06. 008

**引用格式**：张逸勤，杨杰，宋欣雨，邓三鸿，赵星. 基于代表作的科学家学术影响力评价模式研究——以诺贝尔奖得主为例［J］. 情报理论与实践，2025，48（6）：61-68.

**Research on the Evaluation Model of Scientists' Academic Impact Based on Representative Works:**

**A Case Study of Nobel Prize Winners**

Zhang Yiqin[1, 2]，Yang Jie[1, 2]，Song Xinyu[1, 2]，Deng Sanhong[1, 2]，Zhao Xing[3, 4]

（1. School of Information Management，Nanjing University，Jiangsu Nanjing 210023；
2. Key Laboratory of Data Engineering and Knowledge Services in Provincial Universities （Nanjing University），Jiangsu Nanjing 210023；
3. Institute of Big Data（IBD），Fudan University，Shanghai 200433；
4. National Institute of Intelligent Evaluation and Governance，Fudan University，Shanghai 200433）

**Abstract**：［Purpose/significance］Academic evaluation plays a key role in talent selection，project assessment，and science and technology policy formulation. With the deepening of the “breaking the five-only” reform，the representative work evaluation system has garnered increasing attention as a new evaluation model. However，there is currently a lack of systematic quantitative analysis to support it.［Method/process］This study introduces “average citations per representative work” as a quantitative indicator of scientists' academic impact. Drawing from the Microsoft Academic Database （MAG），301 Nobel Prize laureates in Physics，Chemistry，and Physiology or Medicine，alongside 2518 control group scientists，were selected as research subjects. By analyzing the distribution of paper outputs，the top five papers ranked by citation frequency in descending order were identified as representative works. Multiple statistical methods were applied to conduct the empirical analysis.［Result/conclusion］①Compared to traditional bibliometric indicators such as the h-index，the evaluation model based on representative works demonstrates superior discriminant validity. ②The indicator shows strong alignment with expert judgments of Nobel Prize winners. ③The evaluation model displays robust applicability across different academic disciplines. This study provides a novel theoretical framework and empirical basis for the quantitative evaluation of scientists' academic impact，delivering valuable theoretical and practical guidance for academic evaluation system reforms.［Limitations］This study primarily focuses on mature scientists within the natural sciences，and further validation is needed regarding its applicability to the humanities，social sciences，and early-career researchers.

**Keywords**：representative work；Nobel prize；citation analysis；academic evaluation

* 本文为国家社会科学基金重大项目“面向国家战略的情报学教育与发展研究”（项目编号：20&ZD332）和国家自然科学基金青年学生基础研究项目（博士研究生）“融合创新测度和生成模型的科学技术知识空间预测研究”（项目编号：724B2015）的成果。

## 0 引言

学者层级的学术评价是科研评价体系的关键环节[1]，不仅在人才遴选、项目评审和绩效考核等方面[2-3]发挥重要作用，还对机构、地区和国家层面的科技评价具有指导意义[4]。准确评估学者的学术影响力，既关系到科学家个体的发展轨迹[5]，也关系到整个科研生态环境的健康发展[6]。然而，传统的学术影响力计量指标过分强调产出数量而轻视创新质量[7]，其有效性饱受质疑[8]，难以真实反映学者的实质性贡献[9]。在“破五唯”改革背景下，学术评价的重心已转向研究质量。自2016年以来，中共中央办公厅、国务院办公厅等多个国家部门陆续印发了包括《关于深化项目评审、人才评价、机构评估改革的意见》在内的相关政策文件，明确要求推行代表作评价制度，注重标志性成果的质量、贡献与影响。特别是近期科技部等八部门联合发布的《关于开展科技人才评价改革试点的工作方案》和启动的试点工作，以“分类评价”和代表作评价为核心，标志着科技评价改革进入深化阶段。

代表作制度作为新型学术评价模式已进入实践应用阶段，但其实施路径、具体模式及有效性尚缺乏系统的定量研究支撑。为此，本文提出一种简洁、透明且具有可拓展性的代表作评价模式，即以“代表作篇均被引量”作为科学家学术影响力的定量表征。基于微软学术数据库（MAG）[10]，本文采用多元统计方法对物理学、化学和生理学或医学领域的311位诺贝尔奖得主及2518位对照组科学家进行实证分析，以验证该评价模式的有效性。

## 1 问题探讨

论文的引用描绘了科学知识的继承与发展的过程[11]。基于被引量的定量指标[12]作为评价基石，广泛应用于学者评价[13]，其中总被引量、h指数[14]和g指数[15]最为普遍[16]。然而，这些传统指标在较大程度上受限于科学家的发文量，其有效性受到了广泛的质疑[17]。以诺贝尔物理学奖得主著名科学家杨振宁和希格斯为例，如图1所示，两位学者具有相近的学术生涯区间[18]。杨振宁发表了百余篇学术论文，其中包含20余篇高被引论文，h指数为52、g指数为131；相较而言，希格斯虽然第一代表作被引量高达2816，但因发文量较少，其h指数仅为8、g指数仅为16。由此案例可见，h指数和g指数等传统指标在评价科学家代表性成果影响力时存在表征效应不足[19]的局限。近年来，学者们通过《莱顿宣言》[20]和《旧金山宣言》[21]相继提出对传统评价指标的批判，呼吁构建新型评价体系。

在此背景下，代表作评价制度应运而生[22-23]，Bornmann等[24]建议以代表作的数量而非h指数来评价学者；Wang等[25]将学者代表作的被引量和颠覆性纳入评价框架。宋丽萍等[26]探讨了开放科学环境下的评价路径；谢珍等[27]基于文本网络和引文网络，提出了面向代表作评价的一种论文创新性测度方法。杨瑞仙等[28]系统梳理了我国学术代表作评价的研究进展，发现目前代表作评价的研究以理论探讨为主，而缺乏定量的实证支撑。目前，国内采用的“作者自评”与“同行评议”相结合的评价方法，存在主观性较强、效率低下[29]等问题，且容易受到认知偏差[30]、同行选择[31]以及人情关系[32]的影响。因此，建立科学有效的定量指标作为同行评议的辅助依据成为当下亟待解决的问题。

与此同时，由于科学家的学术生涯呈现出显著的阶段性特征，如初期阶段重在潜力展现、中期阶段注重成果积累、成熟期则体现在标志性贡献上。基于代表作的评价模式需要考虑这些不同发展阶段的特点。本文以诺贝尔奖得主为样本，主要针对成熟期科学家群体，其评价重点在于甄别重大学术贡献。对于处于不同发展阶段的科研人员，评价指标的选择和权重配置仍需要进行相应调整。

近年来，得益于大规模学术数据的开放共享[33]，科研人员得以从宏观视角对传统计量指标和新评价模式进行定量比较。基于此，本文整合开放数据源和多种定量指标收敛有效性测度模型[34]，通过对诺贝尔奖得主及其对照组的系统分析，为代表作评价模式的可行性和有效性提供实证依据。

## 2 数据与指标

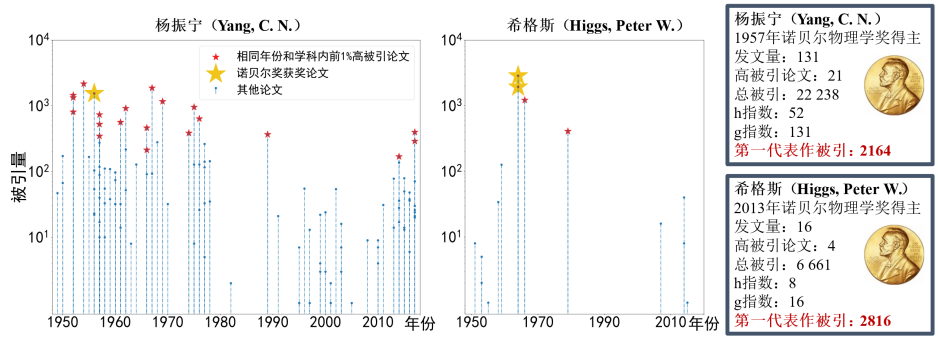



图1 传统评价指标的局限性：以两位著名物理学家为例（引文数据来源为MAG）

Fig. 1 Limitations of traditional evaluation metrics: an example of two famous physicists (citation data source: MAG)

### 2.1 引文数据

MAG数据库是全球最全、最大的学术数据库之一[10]，包含1800—2021年的2亿余篇文献和17亿余条引用链接，以及相应期刊年卷号、标题、DOI等信息。MAG数据库采用了神经网络和众筹法进行了作者名称消歧[35]，共有2亿余位消歧后的作者。此外，本文采用MAG数据库提供的OECD（经济合作

与发展组织）二级分类号作为“分类评价”的依据之一。

### 2.2 诺贝尔奖得主数据

基于Li等[36]提供的物理学、化学和生理学/医学领域543位诺贝尔奖得主数据，进一步筛选出能够在MAG数据库中匹配出其获奖论文期刊年卷号的361位诺贝尔奖得主作为研究样本，其基本信息如表1所示，样本中各个领域的诺贝尔奖得主数量大致相同，平均获奖年份在1973—1978年之间，获奖滞后时间为16.5～18.5年。

表1 诺贝尔奖得主的样本数据描述

Tab. 1 Description of sample data of Nobel laureates

| 领域 | 数量 | 获奖论文平均被引量 | 获奖论文平均发表年份 | 平均获奖年份 | 获奖滞后/年 |
|---|---|---|---|---|---|
| 物理学 | 120 | 2412 | 1955 | 1973 | 18.24 |
| 化学 | 114 | 3196 | 1961 | 1978 | 16.80 |
| 生理学/医学 | 127 | 3277 | 1960 | 1978 | 17.45 |

为定量分析代表作评价模式，考虑到不同学科、不同年份中论文的引用模式差异显著[37]，本文采取“分类评价”和“同时段评价”的方法。其中，“分类评价”通过结合期刊和OECD二级分类号，保证每一位诺贝尔奖得主和对照组中的对应科学家处于同一学科领域。期刊是最细粒度的学科分类指标之一[38]，而OECD二级分类号也在一定程度上解决了综合类期刊的跨学科研究分类问题。

借鉴闵超等[39]选取诺贝尔奖论文对照组的方法，基于以下步骤选取物理学、化学和生理或医学诺贝尔奖得主的对照组。①基于MAG数据库检索诺贝尔奖得主的获奖论文，提取出其年份、标题、期刊、卷、期、学科分类号等信息；②从MAG数据库中选取与诺贝尔奖得主获奖论文在同一年份、同一期、同一卷发表的论文，并抽取出同一学科类别下的论文（OECD二级分类号），作为获奖论文的对照论文集；③从获奖论文的对照论文集中，提取出这些论文的全部作者，并在MAG数据库提取出这些作者的论文信息，计算其发文量，并选取发文量最高的10位科学家加入对照组。

通过以上方式进行抽取，对于每一位诺贝尔奖得主，最终得到2518位对应科学家，作为本文的对照组。对照组的数据基本描述统计如表2所示，各个领域对照组的数量基本与诺贝尔奖得主数量相对应。

诺贝尔奖得主和对照组的论文时间分布如图2所示，可以看出，本文所选取的对照组科学家的学术生涯区间与诺贝尔奖得主基本一致，论文数量的时间分布基本相同。

### 2.3 代表作评价指标

一般认为，代表作是最能够反映一名学者自身研究水准和影响力的一部或几部作品[22, 28]。然而，学界尚未就代表作的选取标准达成一致。被引量被认为是最直观、最简单、最基本和最有效的单篇论文影响力测度指标[39]，在“分类评价”的条件下，被引量能定量反映出代表作的水平和影响力。因此，本文将科学家发表的所有论文按被引频次降序排列，选取排名前$N$篇论文作为其代表作。例如，本文将一名科学家生涯中被引量最高的一篇论文视为其“第一代表作”，并以此类推。对于不同学科的学者，统一选取相同数量的代表作，计算其“篇均被引量”，并以此作为代表作影响力的度量指标用于定量描述该学者的影响力。

表2 对照组样本数据描述

Tab. 2 Description of control group sample data

| 领域 | 数量 | 第一代表作平均被引量 | 平均发文量 | 平均总被引 | 平均h指数 | 平均g指数 |
|---|---|---|---|---|---|---|
| 物理学 | 874 | 1447 | 323 | 13591 | 34 | 66 |
| 化学 | 767 | 1333 | 323 | 16599 | 46 | 84 |
| 生理学/医学 | 877 | 1326 | 213 | 15546 | 43 | 79 |

以学者代表作的“篇均被引量”作为代表作评价的定量测度指标，成为代表作评价制度实施路径的基本组成单元，这种方式不仅能够定量反映学者的代表作影响力，并且具备简单、可行、直观、透明的优势，符合定量评价指标的基本要求[40]，有助于在未来进行大规模地推广和实施。

## 3 结果与讨论

### 3.1 样本均值差异

为系统比较代表作评价指标与传统指标（发文量、总被引、h指数、g指数和“第一代表作”被引量）的区分效果，本文首先对三大领域的诺贝尔奖得主及其对照组进行均值差异分析。由于各项指标不满足同方差性和正态分布假说，本文采用了Welch双尾$t$检验进行组间差异显著性检验。Welch双尾$t$检验是一种非参数检验方法，能够满足本文对组间差异对比的需求。如图3所示，结果揭示了三大领域的诺贝尔奖得主在发文量方面均显著低于对照组（$P$<0.1），这一现象凸显了单纯以产出数量评价科学家

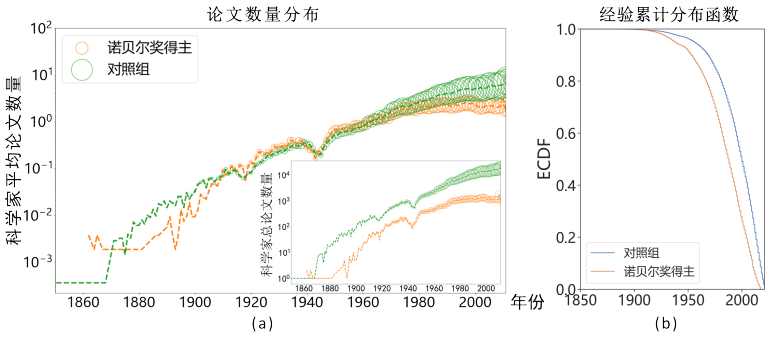



图2 诺贝尔奖得主与对照组的论文数量时间分布

Fig. 2 Time distribution of the number of papers by Nobel laureates and the control group

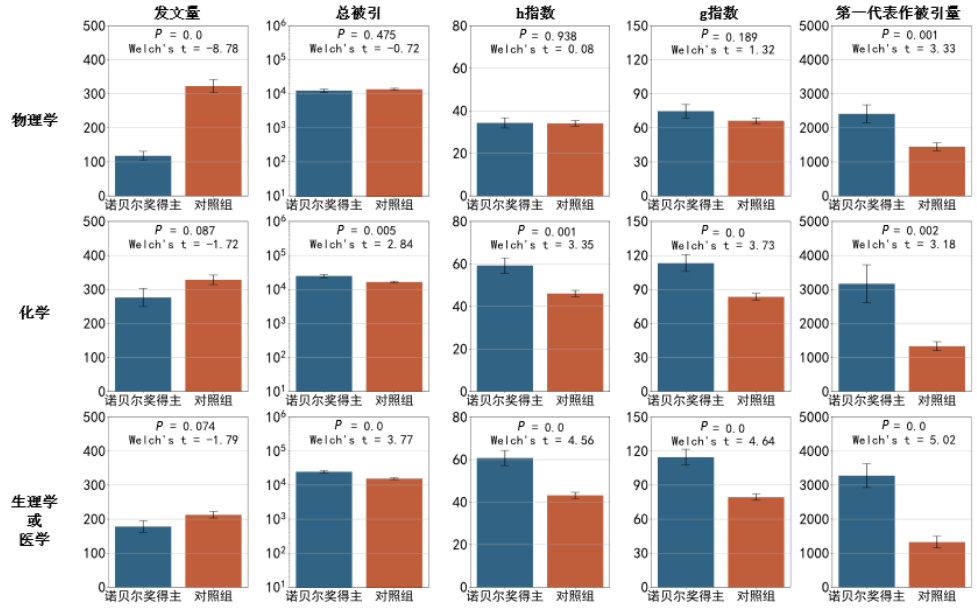



图3 诺贝尔奖得主与对照组的指标均值差异（误差棒：95%置信区间）

Fig. 3 Mean differences of metrics between Nobel laureates and the control group （error bars： 95% confidence interval）

的局限性。同时传统计量指标在不同学科领域表现出显著的不一致性。具体来说，总被引这一指标无法区分诺贝尔物理学奖得主和其对照组（$P$=0.475）；而h指数和g指数这两个指标虽然在化学和生理或医学领域中表现出显著的区分效果（$P$<0.01），但在物理学领域却无显著差异（$P$=0.938或$P$=0.189）。相较而言，“第一代表作被引”指标在三大领域中均呈现出了显著的区分效果（$P$<0.01），表明该指标具有更强的普适性和稳健性。这一结果为代表作评价指标的有效性提供了初步的实证支持。

### 3.2 相关性矩阵

为分析代表作评价指标与传统计量指标的差异性，本文采取肯德尔秩（Kendall’s tau）相关性系数[41]进行相关性分析。选择该系数的原因在于h指数、g指数、发文量和被引量等指标均为可直接用于排序的正整数[41]。相关性分析矩阵如图4所示，结果显示，代表作评价指标在评价诺贝尔奖得主和对照组科学家时呈现出显著不同的相关模式。对于诺贝尔奖得主而言，代表作评价指标之间具有很强的内部一致性（$\tau$>0.83），尤其是“4篇代表作均引”和“5篇代表作均引”的相关性系数高达0.98。然而，代表作评价指标与传统计量指标的相关性较低，其中发文量的相关性较低（$\tau$<0.33），与h指数和g指数的相关性也不高（$\tau$<0.55）。这表明在评价顶尖科学家时，代表作评价指标提供了一个独特的评价维度。相较之下，在对照组中，除发文量以外的传统计量指标均与代表作评价指标高度相关（$\tau$>0.7）。该结果说明，代表作评价指标在测度顶尖学者和其他学者的影响力时，与传统计量指标的相关性存在显著区别。代表作评价指标在评价普通科学家时与传统指标具有较好的一致性，但在识别和评价顶尖科学家时则展现出独特的区分能力，这种差异化特征使其特别适合用于高层次人才评价。

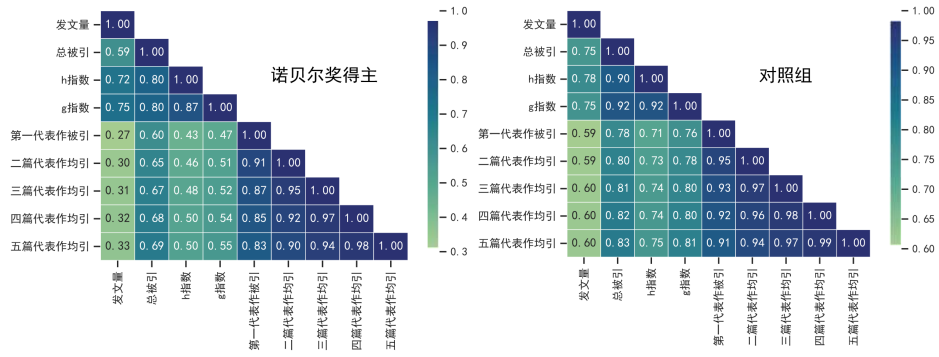



图4 传统计量指标及代表作评价指标相关性矩阵

Fig. 4 Correlation matrix of traditional quantitative metrics and representative work evaluation metrics

### 3.3 收敛有效性

相关性分析说明了代表作评价指标在评价诺贝尔奖学者时与传统计量指标的结果显著不一致。为进一步证明代表作评价指标的收敛有效性，即该指标能够在多大程度上反映学者的真实影响力[42]，本文提出两种评价指标的收敛有效性测度模型：诺贝尔奖得主的识别率（Identification Proportion，IP）和诺贝尔奖得主的平均排名（Average Ranking，AR）。

首先，分学科将诺贝尔奖得主与对照组科学家进行合并，针对某个特定的评价指标，IP表示以该指标识别出的排名最高的前$n$%学者中的诺贝尔奖得主占比。识别率IP与机器学习中的召回率（Recall）理念一致，IP越高，说明指标的收敛有效性越强[43]，本文将平均IP定义为阈值$n$取1%～100%（以1%为间隔）之间的IP平均

值，并以平均IP作为衡量指标有效性的主要参数。

平均排名AR定义为该指标下的全部诺贝尔奖得主的平均排名，排名越靠前表明指标越有效。由于简单地取均值会受到其他学者排名的影响，除基准AR以外，借鉴Mariani等[44]的研究思路，取排位最低的诺贝尔奖得主和排位最高的诺贝尔奖得主的排名之差作为分母，得到标准化的平均排名作为鲁棒性检验。IP主要从召回率的角度测量了指标识别出的重要学者中诺贝尔奖得主的比例，AR则从全部诺贝尔奖得主的平均排位出发，它们都能够定量反映出计量指标的收敛有效性[45]。

评价指标收敛有效性的结果如图5所示，在物理学、化学和生理学或医学这三大领域，代表作评价指标的收敛有效性均明显高于发文量、h指数和g指数等传统指标。具体而言，代表作评价指标在不同阈值下的诺贝尔奖得主识别率均显著高于传统指标，对诺贝尔奖学者的平均排名和标准化排名均为最小，这表明代表作评价指标能够更为有效地识别出诺贝尔奖得主。

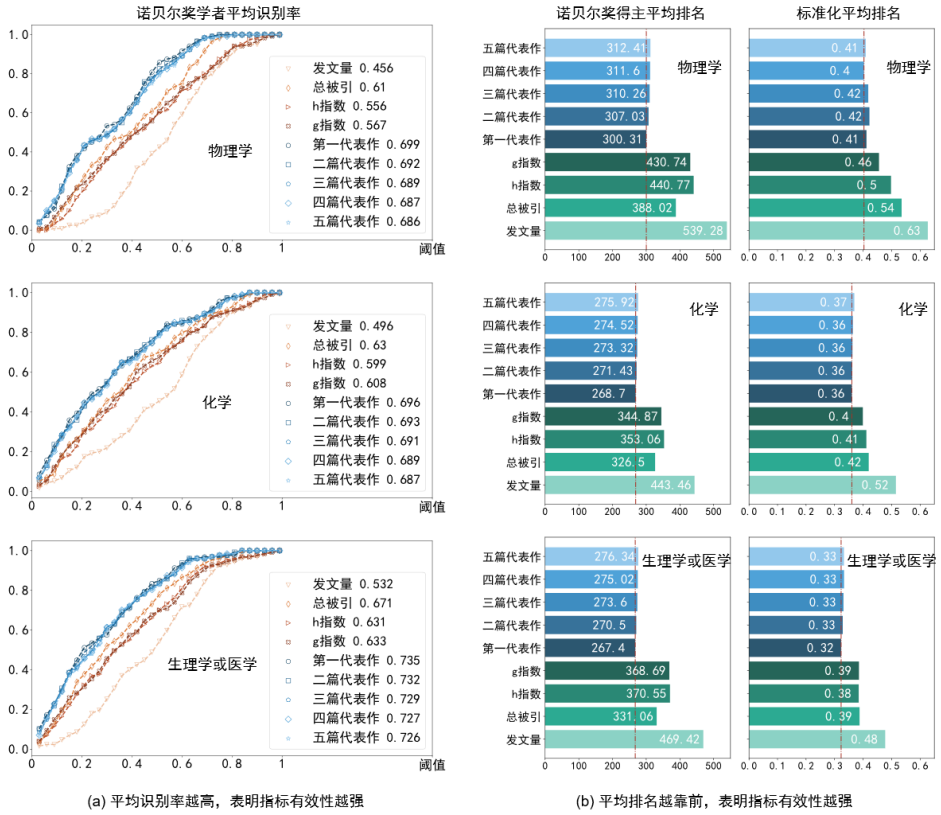
值得注意的是，5篇以内的“代表作均引”的收敛有效性都显著优于传统指标，并且不同数量的代表作平均指标的有效性差异并不明显，但是从总体来看，“第一代表作被引”的有效性是所有评价指标中最高的。它在物理学、化学和生理学或医学的平均识别率IP分别为0.182、0.216和0.229，在识别顶尖学者任务的识别率甚至达到了0.4以上（图5（a），生理学或医学领域，阈值$n$=0.02）。从平均排名AR的结果来看（图5（b）），除了在标准化平均排名这一任务上弱于“4篇代表作均引”以外，“第一代表作被引”下的诺贝尔奖得主平均排名基本都靠前，也佐证了“第一代表作被引”的有效性最强这一结论。尽管如此，其他“代表作均引”指标在一些细分任务上出现了有效性高于“第一代表作被引”的情况，如特定学科和阈值下的识别率。

这一结论具有重要的启示和意义，首先它证明了在本文的视角下，代表作评价指标作为学者的影响力测度，相比传统的计量指标具备全面的优越性。其次，该结论还表明在本文探讨的所有评价指标中，“第一代表作被引”的总体有效性是最高的，但是它在一些细分任务上可能弱于其他“代表作均引”指标。这或许反映了学者评价的一个最优路径：以“第一代表作被引”作为主要影响力测度指标，以其他数量的“代表作均引”作为替代与辅助，从而解决代表作评价模式的量化问题和具体路径实现。

### 3.4 逻辑回归

为进一步证明代表作评价指标的有效性，本文以“是否为诺贝尔奖得主”（二元分类）为因变量，构建逻辑回归模型以检验评价指标对诺贝尔奖学者的预测效应。为克服不同个体的外生性差异影响，基于前文的对照组科学家匹配方式构建个体随机效应模型[46]。

由于传统评价指标内部和代表作评价指标内部存在很强的多重共线性，基于VIF检验结果进行变量剔除，最终对于传统评价指标和代表作评价指标仅保留一种，构建4种模型：发文量和“代表作均引”（模型1），总被引和“代表作均引”（模型2），h指数和“代表作均引”（模型3），g指数与“代表作均引”（模型4）。4个模型中的VIF值均小于2，预先对各个自变量进行0-1标准化后，对“是否为诺贝尔奖得主”预测变量进行分学科回归，结果如图6所示（图中仅呈现“第一代表作被引”，研究团队对其他“代表作均引”指标也进行了回归，结果基本一致），红色




(a) 平均识别率越高，表明指标有效性越强　　(b) 平均排名越靠前，表明指标有效性越强

图5　评价指标的收敛有效性：诺贝尔奖得主的识别率与平均排名

Fig. 5　Convergence validity of evaluation metrics： identification rate and average ranking of Nobel laureates

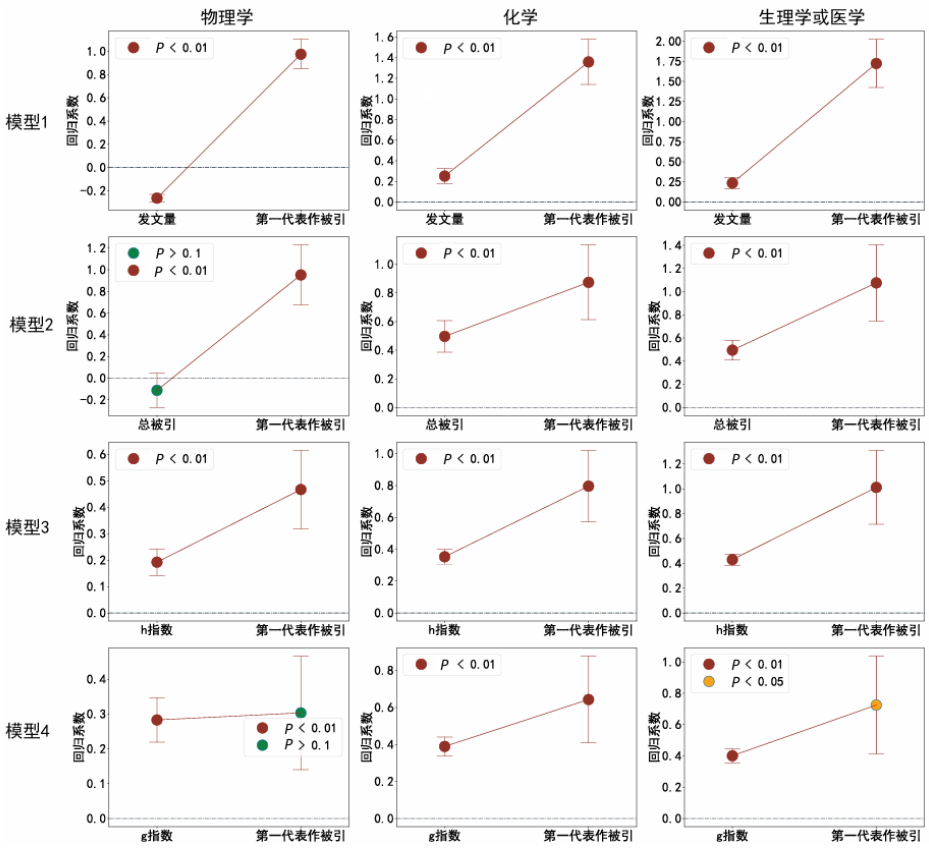



图6 逻辑回归结果（个体随机效应模型，误差棒：95%置信区间）

Fig. 6 Results of logistic regression （individual random effects model， error bars： 95% confidence interval）

点（计算机显示颜色，下同）表示在0.01水平下显著，橙色点表示在0.05水平下显著，绿色点则表示不显著。

显然，代表作评价指标在物理学、化学和生理学或医学这三大领域中以及各种回归模型中的回归系数基本上都是显著的正值，尽管在模型4中，物理学、生理学或医学领域中的代表作评价指标显著性不如g指数，但是它的回归系数更高，因此没有足够的证据能够表明代表作评价指标的预测效应弱于传统指标。

反观传统计量指标中，发文量和总被引在物理学领域的回归系数为负值，在一定程度上表明仅凭发文量和总被引难以识别出该学科的顶尖学者。在其余模型以及各个领域中，传统计量指标的预测效应都明显低于代表作评价指标。这一结论充分验证了上述分析结果的鲁棒性，进一步证明了本文提出的代表作评价指标相对于传统计量指标的优越性和有效性。

## 4 总结与未来展望

### 4.1 总结

本文从科学家学术影响力评价的实证视角，构建并验证了基于代表作的量化评价模式。通过对301位诺贝尔奖得主及其对照组的系统分析，研究主要获得以下发现：首先，传统计量指标在评价高影响力科学家时存在明显局限性，而“代表作篇均被引”指标展现出更高的评价效度；其次，该评价模式在物理学、化学和生理学或医学三大领域中均呈现出显著的优越性和稳健性，与传统h指数等计量指标的低相关性（$\tau$<0.55），表明其提供了独特的评价维度；最后，该指标在识别高影响力科学家方面具有更强的区分能力，其对诺奖得主的识别准确率显著高于传统指标。

从理论价值来看，本文所构建的“代表作篇均被引”新型评价模式不仅具备计算简便、过程透明的特点，而且在评价科学家实质性学术贡献方面显著优于传统计量指标。研究发现该指标与专家评价结果具有高度一致性[45]，这表明其可作为同行评议的重要补充，为建立更科学的学术评价体系提供了新思路。从实践层面看，该代表作评价模式有助于引导科研人员更加注重成果的质量与长期影响力，而非短期的数量产出，有效避免科研领域中出现“快餐式科研”或“重数量轻质量”等不良现象。同时，该评价体系也可为人才遴选、项目评审等工作提供客观依据，以期更加公正、客观、有效地评估科技人才和科技成果的价值和贡献。

### 4.2 未来展望

值得注意的是，代表作评价制度及其指标的应用也面临着两个主要挑战。首要挑战是学科差异带来的适用性问题。不同学科领域的引用模式存在显著差异[37]，如本文涉及的物理学、化学与生理学或医学领域在发文量和被引特征上均表现出较大的学科差异性。这一发现表明，在实施代表作评价时必须坚持“分类评价”原则，将评价对象限定在同一学科范围内进行横向比较，避免跨学科的直接对比。第二个挑战在于评价对象的局限性。本文的实证分析主要基于已取得重要成就的成熟科学家群体，这使得评价模式在应用于不同职业发展阶段的科研人员时可能存在偏差。特别是对于青年学者而言，由于其学术影响力正处于积累阶段，现有的评价方案可能无法准确反映其学术潜力，需要采用更灵活的选择标准和合理的评价周期。

基于本文的发现与局限，未来研究可从以下4个方向

深化和拓展：第一，扩大评价模式的适用范围，重点探索其在人文社科等领域的应用价值。当前研究仅限于自然科学领域，未来需要结合人文社科的学科特点，整合基金项目、重要奖项等多源数据，建立更具学科特色的评价标准体系。第二，构建更加精细的贡献度量化方法，重点关注科研合作维度。随着科研合作日益普遍[47]，有必要将作者署名顺序[40]、团队协作模式等因素纳入评价框架，以准确反映个体学者在团队成果中的实质贡献[48]。第三，拓展学术评价指标的多维性。除被引量[7]核心指标外，建议引入下载量[49]、学术社交网络影响力、政策引用[4]等替代计量指标[50]，同时考虑对学术文本内容的深层分析。通过指标的标准化处理[51]和多维整合，有助于构建更全面的影响力评估模型。第四，建立动态评价机制。针对不同职业发展阶段，特别是青年学者群体，需要开发更具预测性的评价方案。通过构建差异化的评价指标体系，将代表作评价与学术潜力预测相结合，开发能够动态捕捉学者成长轨迹的评价模型。这些深化研究将有助于使代表作评价模式更加科学完善，从而更好地服务于学术评价的公平性和有效性。

本研究支撑数据：①微软学术数据库（MAG）（https://learn. microsoft. com/en-us/academic-services/graph/get-started-setup-provisioning）；②诺贝尔奖得主论文数据（https://www.nature.com/articles/s41597-019-0033-6#MOESM1）。□

**参考文献**

[1] HOFSTRA B，KULKARNI V V，MUNOZ-NAJAR GALVEZ S，et al. The diversity-innovation paradox in science [J]. Proceedings of the National Academy of Sciences，2020，117（17）：9284-9291.

[2] DING Ying，CRONIN B. Popular and/or prestigious? Measures of scholarly esteem [J]. Information Processing & Management，2011，47（1）：80-96.

[3] YIN Yian，WANG Yang，EVANS J A，et al. Quantifying the dynamics of failure across science，startups and security [J]. Nature，2019，575（7781）：190-194.

[4] YIN Yian，GAO Jian，JONES B F，et al. Coevolution of policy and science during the pandemic [J]. Science，2021，371（6525）：128-130.

[5] DOROGOVTSEV S N，MENDES J F F. Ranking scientists [J]. Nature Physics，2015，11（11）：882-883.

[6] INIGUEZ G，PINEDA C，GERSHENSON C，et al. Dynamics of ranking [J]. Nature Communications，2022，13（1）：1646.

[7] KWON D. The rise of citational justice [J]. Nature，2022，603（7902）：568-571.

[8] WALTMAN L，VAN ECK N J. The Inconsistency of the h-index [J]. Journal of the American Society for Information Science and Technology，2012，63（2）：406-415.

[9] AZOULAY P，GRAFF-ZIVIN J，UZZI B，et al. Toward a more scientific science [J]. Science，2018，361（6408）：1194-1197.

[10] WANG Kuansan，SHEN Zhihong，HUANG Chiyuan，et al. Microsoft academic graph：when experts are not enough [J]. Quantitative Science Studies，2020，1（1）：396-413.

[11] FORTUNATO S，BERGSTROM C T，BOERNER K，et al. Science of science [J]. Science，2018，359（6379）：eaao0185.

[12] BELTER C W. A relevance ranking method for citation-based search results [J]. Scientometrics，2017，112（2）：731-746.

[13] XIA Wanjun，LI Tianrui，LI Chongshou. A review of scientific impact prediction：tasks，features and methods [J]. Scientometrics，2023，128（1）：543-585.

[14] HIRSCH J E. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy of Sciences，2005，102（46）：16569-16572.

[15] EGGHE L. Theory and practise of the g-index [J]. Scientometrics，2006，69（1）：131-152.

[16] ZHANG Yu，WANG Min，ZIPPERLE M，et al. RelRank：a relevance-based author ranking algorithm for individual publication venues [J]. Information Processing & Management，2023，60（1）：103156.

[17] 朱军文，刘念才. 高校科研评价定量方法与质量导向的偏离及治理 [J]. 教育研究，2014，35（8）：52-59.（ZHU Junwen，LIU Niancai. The deviation of quantitative evaluation of university research performance from quality-orientation and its effective governance [J]. Educational Research，2014，35（8）：52-59.）

[18] LIU Lu，WANG Yang，SINATRA R，et al. Hot streaks in artistic，cultural，and scientific careers [J]. Nature，2018，559（7714）：396-399.

[19] BRITO R，NAVARRO A R. The inconsistency of h-index：a mathematical analysis [J]. Journal of Informetrics，2021，15（1）：101106.

[20] HICKS D，WOUTERS P，WALTMAN L，et al. The Leiden Manifesto for research metrics [J]. Nature，2015，520（7548）：429-431.

[21] CAGAN R. The San Francisco declaration on research assessment [J]. Disease Models & Mechanisms，2013，6（4）：869-870.

[22] NIU Qikai，ZHOU Jianlin，ZENG An，et al. Which publication is your representative work? [J]. Journal of Informetrics，2016，10（3）：842-853.

[23] IOANNIDIS J P A，BOYACK K W，SMALL H，et al. Bibliometrics：is your most cited work your best? [J]. Nature，2014，514（7524）：561-562.

[24] BORNMANN L，LEYDESDORFF L. Count highly-cited papers instead of papers with h citations：use normalized citation counts and compare "like with like"! [J]. Scientometrics，2018，115（2）：1119-1123.

[25] WANG Ruijie，ZHOU Yuhao，ZENG An. Evaluating scientists by citation and disruption of their representative works [J]. Scientometrics，2023，128（3）：1689-1710.

[26] 宋丽萍，王建芳. 开放科学环境下负责任评价实现路径研究 [J]. 中国图书馆学报，2024，50（2）：115-127..（SONG Liping，WANG Jianfang. Study of the realization path of responsible research assessment in the context of open science [J]. Journal of Library Science in China，2024，50（2）：115-127.）

[27] 谢珍，马建霞，胡文静. 面向代表作评价的学术论文创新性

测度方法[J]. 情报理论与实践，2022，45（7）：81-88.（XIE Zhen，MA Jianxia，HU Wenjing. The innovation measurement method for academic papers oriented to the evaluation of representative works[J]. Information Studies：Theory & Application，2022，45（7）：81-88.）

[28] 杨瑞仙，高鑫宁，董克. 我国学术代表作评价研究进展[J]. 图书情报工作，2022，66（17）：129-140.（YANG Ruixian，GAO Xinyu，DONG Ke. Research progress on evaluation of academic representative works in China[J]. Library and Information Service，2022，66（17）：129-140.）

[29] LI Danielle，AGHA L. Big names or big ideas：do peer-review panels select the best science proposals?[J]. Science，2015，348（6233）：434-438.

[30] LEE C J，SUGIMOTO C R，ZHANG G，et al. Bias in peer review[J]. Journal of the American Society for Information Science and Technology，2013，64（1）：2-17.

[31] SILER K，LEE K，BERO L. Measuring the effectiveness of scientific gatekeeping[J]. Proceedings of the National Academy of Sciences，2015，112（2）：360-365.

[32] 俞立平，张矿伟，蒋长兵. 推进代表作评价存在的问题与对策研究[J]. 情报学报，2021，40（4）：345-353.（YU Liping，ZHANG Kuangwei，JIANG Changbing. Promoting the evaluation of representative works：challenges and recommendations[J]. Journal of the China Society for Scientific and Technical Information，2021，40（4）：345-353.）

[33] VISSER M，VAN ECK N J，WALTMAN L. Large-scale comparison of bibliographic data sources：Scopus，Web of Science，Dimensions，Crossref，and Microsoft Academic[J]. Quantitative Science Studies，2021，2（1）：20-41.

[34] BORNMANN L，DEVARAKONDA S，TEKLES A，et al. Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers[J]. Quantitative Science Studies，2020，1（3）：1242-1259.

[35] DORSCH I. Becoming metric-wise：a bibliometric guide for researchers[J]. Journal of the Association for Information Science and Technology，2019，70（5）：530-532.

[36] LI Jichao，YIN Yian，FORTUNATO S，et al. A dataset of publication records for Nobel laureates[J]. Scientific Data，2019，6（1）：33.

[37] ZENG An，SHEN Zhesi，ZHOU Jianlin，et al. The science of science：from the perspective of complex systems[J]. Physics Reports-Review Section of Physics Letters，2017，714：1-73.

[38] UZZI B，MUKHERJEE S，STRINGER M，et al. Atypical combinations and scientific impact[J]. Science，2013，342（6157）：468-472.

[39] MIN Chao，BU Yi，WU Ding，et al. Identifying citation patterns of scientific breakthroughs：a perspective of dynamic citation process[J]. Information Processing & Management，2021，58（1）：102428.

[40] WALTMAN L. A review of the literature on citation impact indicators[J]. Journal of Informetrics，2016，10（2）：365-391.

[41] LÜ Linyuan，ZHOU Tao，ZHANG Qianming，et al. The h-index of a network node and its relation to degree and coreness[J]. Nature Communications，2016，7（1）：10168.

[42] BORNMANN L，TEKLES A. Convergent validity of several indicators measuring disruptiveness with milestone assignments to physics papers by experts[J]. Journal of Informetrics，2021，15（3）：101159.

[43] XU Shuqi，MARIANI M S，LÜ Linyuan，et al. Unbiased evaluation of ranking metrics reveals consistent performance in science and technology citation data[J]. Journal of Informetrics，2020，14（1）：101005.

[44] MARIANI M S，MEDO M，ZHANG Yichen. Identification of milestone papers through time-balanced network centrality[J]. Journal of Informetrics，2016，10（4）：1207-1223.

[45] WANG Jingjing，XU Shuqi，MARIANI M S，et al. The local structure of citation networks uncovers expert-selected milestone papers[J]. Journal of Informetrics，2021，15（4）：101220.

[46] SHOR B，BAFUMI J，KEELE L，et al. A bayesian multilevel modeling approach to time-series cross-sectional data[J]. Political Analysis，2007，15（2）：165-181.

[47] HEIDI L. Team science[J]. Nature，2015，525（7569）：308-311.

[48] SHEN H-W，BARABáSI A-L. Collective credit allocation in science[J]. Proceedings of the National Academy of Sciences，2014，111（34）：12325-12330.

[49] 赵星. 学术文献用量级数据Usage的测度特性研究[J]. 中国图书馆学报，2017，43（3）：44-57.（ZHAO Xing. Exploring the features of usage data for academic literatures[J]. Journal of Library Science in China，2017，43（3）：44-57.）

[50] 赵蓉英，郭凤娇，谭洁. 基于Altmetrics的学术论文影响力评价研究——以汉语言文学学科为例[J]. 中国图书馆学报，2016，42（1）：96-108.（ZHAO Rongying，GUO Fengjiao，TAN Jie. Evaluation of academic papers impact based on Altmetrics：a case study of Chinese language and literature[J]. Journal of Library Science in China，2016，42（1）：96-108.）

[51] WALTMAN L，VAN ECK N J. Source normalized indicators of citation impact：an overview of different approaches and an empirical comparison[J]. Scientometrics，2013，96（3）：699-716.

**作者简介：张逸勤**（ORCID：0000-0002-6499-3115），博士生。研究方向：文本挖掘与数字人文。**杨杰**（ORCID：0000-0002-1276-0483），博士生。研究方向：科学评价与计量。**宋欣雨**（ORCID：0000-0002-6441-7732），博士生。研究方向：科学计量。**邓三鸿**（ORCID：0000-0002-6910-3935，通信作者，Email：sanhong@nju.edu.cn），博士，教授，博士生导师。研究方向：科学评价与信息检索。**赵星**（ORCID：0000-0001-9347-590X），博士，教授。研究方向：科学评价。

**作者贡献声明：张逸勤**，分析数据，修改定稿。**杨杰**，分析数据，撰写初稿。**宋欣雨**，采集、清洗数据。**邓三鸿**，提出写作思路，拟定提纲。**赵星**，参与提纲讨论。

**录用日期：**2025-02-05