

学术评价应更科学

董尔丹*, 胡海, 张俊

国家自然科学基金委员会医学科学部, 北京 100085

* 联系人, E-mail: donged@nsfc.gov.cn

2013-11-11 收稿, 2013-11-12 接受, 2013-12-19 网络版发表

摘要 学术评价是以科学的标准衡量学术活动结果的价值。本文探讨了学术评价的要义和应该遵循的规则, 介绍了学术评价的两种基本方式, 即基于科学内容的学术评价和基于形式的学术评价, 以及两种方式各自的特点及优缺点。综述了学术评价在欧美国家的现状及其面临的问题, 分析了我国现有学术评价方式和存在的问题, 提出了从基于形式为主的评价方式向基于科学内容为主的评价方式转变过程中应该遵循的一些规则和相应的制度建设问题, 并就我国医学学术评价方式的转变和改进提出了建议。

关键词

学术评价
同行评议
科学价值
科学共同体

“科学技术是第一生产力”。新中国成立 60 多年来, 特别是改革开放 30 多年来, 我国的科学技术水平取得了突飞猛进的发展。2002~2012 年我国科技人员共发表各类学术论文 102.26 万篇, 位居世界第 2 位, 被引用次数 665.34 万次, 列第 6 位^[1]。截至 2012 年底, 我国发明专利累计授权量达 111.1 万件, 2012 年全球申请了 210 多万件新专利, 其中中国 65 万件, 约占 30%^[2]。然而需要注意的是, 在快速发展的同时, 出现了一些新情况、新问题, 其中较为突出的是我国科学研究的学术评价方式已经不能适应我国科学技术的快速发展。改革开放早期, 为尽可能避免和减少人为干扰因素, 最大可能体现公平、公正和公开原则, 各大学和科研机构普遍倾向于采用量化指标进行学术评判。其中, 期刊科学引文索引(scientific citation index, SCI)论文及其发表杂志的影响因子(journal impact factor, JIF)是被最普遍应用的指标。当然, 当时采用这一指标的背后还有我国与西方发达国家巨大的科学技术水平差距的无奈, 表现在我们科学界当时缺乏足够的力量去全面评价一些具体领域里科学研究的实际水平。随着我国整体科学研究水平的提升和国际交流的日益广泛, 目前这些情况已基本

不存在。“唯 SCI 和 JIF”的学术评价方式日益凸显其偏颇性和狭隘性, 并在引导我国科技发展的方向上产生了一定程度的偏移。因此, 本文将就学术评价的目标、意义以及相关制度建设作一些探讨。

1 学术评价需适应科学研究的目标并服从科学研究的基本属性

学术评价对于考核学术质量、评价学术价值、鉴定学术成果、引导学术方向具有重要意义。实际上学术评价的意义已远远超越学术的范畴, 还涉及职称评定、基金申请、奖项获取、基地评审、研究资源再分配等诸多问题。因此, 建立科学、合理、准确的学术评价制度, 对于提高我国科学研究的创新能力和竞争力、引导正确的发展方向和良好的学术氛围至关重要。

什么是学术评价? 学术评价是以科学的标准为尺度对学术活动结果作出价值判断^[3]。科学的标准其实内在地包含着科学研究的目的。科学是以世间万象为基础, 用系统的方法对知识的追求, 以及这个过程形成的对自然、社会活动规律的系统认识。也就是说“科学”这个概念本身不仅包括获得新知识的活动, 而且还包括这个活动的结果。按研究对象的不同可

引用格式: 董尔丹, 胡海, 张俊. 学术评价应更科学. 科学通报, 2014, 59: 96-106

Dong E D, Hu H, Zhang J. To improve the research assessment process in China (in Chinese). Chin Sci Bull (Chin Ver), 2014, 59: 96-106, doi: 10.1360/972013-1245

分为自然科学、社会科学、形式科学等。现代科学研究中还包括以人类思维存在为对象的思维科学。按研究结果分类则可分为理论科学和应用科学。简言之,科学研究是追求真理和知识创新的活动,并以其研究成果提升社会实践为目的,这自然构成了学术评价的基础、逻辑起点和最终理由^[3]。

1.1 学术评价要服从科学的社会属性

当今社会,科学研究已成为一门职业。到2010年,我国科技从业人员达354.2万,全世界科技从业人员超过千万^[4]。科学研究早已不是科学家自娱自乐或闭门造车了,科学研究具有明确的社会属性,即科学研究不是科学家个人的事业,而是整个科学界共同的事业。在每一个领域都有众多专门从事该项研究的科学技术人员,每个领域自成系统,同时又是一个更大系统的有机组成部分。在做学术评价时,首先应该站在整个领域的高度,认真考虑这项工作对相关领域整个知识体系的贡献。当然,对知识体系的贡献不仅仅是某个学术问题的阐明,也包括科学现象的发现,科学规律的总结,以及公共数据库的建立,某些专门软件的开发,应用技术的研发,知识转化为产品等等。对整个知识体系贡献价值的评估有时包含着对这些理论、学说、观点、方法的学术价值或理论价值的预测。Francis Bacon在议论学术价值时提出了内部标准和外部标准^[5]。前者是推理和证明,后者是实际应用。同行学术评价所采用的是内部标准,它对基础研究成果的评价比较有效。但Bacon即使对基础研究成果的评价也更看重外部标准,他认为“只有把新概念成功地运用于实际,才是正确性的最终象征”。因此,在进行学术评价时,还要评估相关研究促进知识转化为实际应用所产生的社会价值。

1.2 学术评价要尊重科学研究的自由特性

科学研究以追求真理和知识创新为首要目的。一般来说,思维越自由,思想便越活跃,知识创新的潜力就越大。自由研究是科学的必然属性,科学的繁荣和发展离不开自由的学术精神。复旦大学俞吾金教授把学术活动分为“学术研究活动”和“非研究性的学术活动”两种类型,前者如学术研究课题的酝酿和提出、学术研究的具体过程、学术讨论的展开和深入、学术成果的言说、发表或出版等;后者如学术课题的申报和评审、学术成果的鉴定和评奖、学术组织

的建立和相应的学术领导机构的诞生等^[6]。科学家在进行学术研究活动时应提倡自由、践行自由。这种自由是建立在科学家对整个领域的理解上,做那些至少自己认为是重要的、正待阐明的科学问题。严肃的科学家的的工作应该是系统的,应该有自己专注的领域。进行学术评价时,要充分尊重学术自由。一般来说,进行同行评议之前需要制定共同的价值标准,这种价值标准既不能掺杂非学术意志,也不能有为某种学术观点背书的倾向。但由于每个评价者的思想观点和观察角度不同,评价结论又涉及把科学价值量化的问题,同时评价者通常以自己的科学价值观为参照,因此评价者的观点有时会使同行评议出现偏差。美国曾开展过一项研究,把一篇虚构的关于儿童行为矫正的论文,寄给75位对该问题有明确看法的心理学家进行评议。结果当论文结论符合审稿人的观点时,审稿人一般都打分很高,反之则很低^[7]。另外,学术同行由于经过类似的科学训练,有相近的科学价值观,往往会有趋同的科学思想和研究范式,这有利于学识评价达成共识,但却不利于创新,有时可能会排斥某些创新观点,拒绝承认其价值。因此,在进行学术评价时,除同行评议之外,还需要引入一些客观标准。比如,科研成果发表的论文档次,论文的引用情况等。这些客观标准实际是更大范围内的同行评议,从而避免小范围同行评议的主观意志对科学价值的误判,干扰被评议研究者的学术自由。

1.3 学术评价要注重科学研究的公益特性

美国著名的教育心理学家Lee Shulman提到:并不是所有的智力活动都能称得上学术,一种智力活动要称得上学术,至少必须具备以下条件:它必须是公开的,必须接受学术团体中其他成员的批判性评价,学术团体中的所有成员要能利用、并能促进和发展这些思想和实践。也就是说,它必须成为一种“共同的财富”^[8]。很显然,要满足上诉要求则离不开学术的另外一个基本特征——公益性。如果严格的按Shulman的定义,则密而不宣的研究成果就不能作为一种思想或创造活动在学术界中传播、使用、参考和发展,不能对人类知识体系作出贡献,也就不能成为学术评价对象。美国贝勒大学教师教育学院的学术评价标准规定为“学术创造知识或以一种新的方式使用现在的知识解决疑难、回答问题”^[8]。也就是说除创造知识外,对已有的学术成果的综合和再解释也是

学术活动,这种学术活动产生的一般是论著、综述等成果.学术成果是和学术职称密切挂钩的,一般的学术职称包括助理教授、副教授、教授和院士等.其他如工程师、医师等应该不能算严格意义的学术职称.一般来说,有了学术成果的人才能拥有学术职称,并统称为科学家.

2 学术评价的方式和原则

学术评价一般有两种方式:一是基于科学内容的学术评价,即根据科学发展的内在规律和学科本身的逻辑结构进行评价.这种方法要求评审人注意集中在学科内部,实质性地了解科学研究的内容后对其科学价值作出评价.基于内容的评价要求评价人能够读懂和理解学术成果的内容,对该学术领域有深刻的了解和非凡的洞察力.这需要评价人对近期这个领域里的研究情况有充分的了解和总结,分析出近期本学科的重要进展和未来发展趋势.因此,评价人通常是同领域研究者,即同行,最好还是领先的研究者.这种方法往往受到评议人主观和客观因素的影响,如评议人的知识结构、学术兴趣、道德心理、人际关系、工作责任心等.为了尽量降低这些因素的负面影响,基于内容的评价往往需要有多人评价人,并要求他们从规定的不同方面进行评价.二是基于形式的评价方法.形式化评价实质是基于内容评价方法的外化.任何学术成果都有其外在表现特征,评价者通过描述学术研究成果的外在特征来反映学术成果的科学价值.形式化评价考察的是研究成果以哪些形式呈现出来,包括发表文章的数量和杂志,文章被引用的情况;申请专利和授权专利的数量,专利转让情况;研究成果已经和预期产生的经济效益、社会效益等.

1955年,美国学者 Eugene Garfield 博士首次提出可通过论文的被引证频次来衡量期刊的影响^[10];1961年,在美国国立卫生研究院(NIH)支持下,科学引文索引(SCI)得以出版;1972年,科学杂志的影响因子(JIF)被作为期刊评价的工具,根据期刊所发表论文被引用的频次和影响来划分期刊等级,从而使期刊的定量评价成为可能^[11].

IF 是一个比值,其计算公式为:

$$IF = \frac{\text{某期刊前两年发表的文献在评价当年被ISI收录期刊所引用的总次数}}{\text{前两年该期刊出版的可引用文献总数}}$$

注: ISI—Institute of Science Information, 美国科学情报研究所

但 SCI 和 JIF 有其明显的局限性.一个杂志的 IF 不能和其发表文章的科学水平划等号.研究表明一个杂志发表的约 85% 的文章引用次数实际低于该杂志文章被引用的平均水平^[12].一个极端的例子是发表于 *Acta Crystallographica Section A* 的一篇论文“*A short history of SHELX*”,该论文发表后被引用 6600 余次,导致该期刊的 IF 从 2008 年的 2.051 飙升至 2009 年的 49.926.而在该期刊发表的所有论文中,引用频次居第二位者仅被引用 28 次.另外,IF 反映的是期刊最近 2 年发表论文平均被引频率,但是 Krell 等人^[13]研究发现,98.5% 被引用文献的发表年份在计算期限的 2 年之后.此外,即使只考察文章引用次数也存在明显的问题,因为有些是作者自己或朋友间的引用,有些引用并不是对文章阐述主要科学问题的支持.显然简单化的仅用 JIF 作为量化学术评价的依据是不科学的,甚至是粗暴的.因此, Eugene Garfield 于 1998 年发表了题为“*The impact factor and using it correctly*”的文章,指出 IF 正在被滥用,同一期刊发表的论文,其科学价值可能千差万别,把对期刊影响力的评价指标扩展为对研究论文的评价显然是不合理的^[14],如果进一步扩展为对科研成果、科研团队和科学家的评价指标则无疑是过度滥用了.

总体来说,这两种评价方法是相辅相成的.不同的科学发展阶段可能会侧重于某种方法,但基于内容的评价显然更符合科学的内在逻辑.我国在改革开放早期普遍采用形式化的评价方式.客观上说,这种方法在初期对推动我国科学水平的发展起到了积极的作用.然而鉴于形式化评价产生的负面影响,近年来中国科学界再度响起了学术评价回归根据科学内容进行的同行评议的呼声.由于同行评价也存在前面提出的干扰因素,为减少其负面影响,科学界也为同行评价制定了一些操作的规则.

第一,学术评价的民主规则.任何评价人的意见在学术评价中的权重都应相等.当然,这种民主只是参与评价的人的民主.没有参与这一过程的人也可能做出更加准确的评议.

第二,评价过程的单盲规则.单盲是指被评价人不知道有谁参与了学术评价,而评价人未经允许也不能公开其评价人的身份,以最大限度地减少评价人从评价中谋利的可能性.一般来说,在通讯评审中,回

避规则的实施比会议评审更加容易和彻底。有人提出双盲的评审方式,但实际上并无可操作性,因为很多时候,评价人根据评议的材料即可推断出被评议人。

第三,评价机构中立规则。学术评价的权威性来源于评价机构的中立性,即不受任何非学术因素干扰,包括行政干预、评价对象或利益集团的劫持等。

第四,评价人分散回避规则。为了减少评价人的人际关系等客观因素对其评议的影响,一般要求本人回避,即申请者不能任评委。另外还要求利害关系人回避,即与评价对象有利害关系者回避。

需要指出的是,同行评议尽管有上述的一些操作原则,现实中即使严格执行也还是会有一些问题,毕竟评价的结果只代表少数人的意见。用少数人的判断来裁决学术价值的有无或学术水平的高低,有时会造成荒谬的结果。John Rawl 就对程序正义的绝对化有保留,他说:“不能因为一种特殊结果是在遵循一种公平的程序中达到的就说它是正义的。这个口子开得太大,会导致荒唐的不公正的结果”^[15]。

3 欧美国家的学术评价情况

总体来说,欧美国家倾向于同行评议。比如哈佛大学规定,在晋升正副教授时,所在院系部门需组织专家委员会对候选人的学术影响力进行评估,论文发表的数量和 JIF 不是主要参考指标,并未对论文发表数量和级别有明文硬性规定。“是否被国内同行认可可是晋升副教授的标准,是否被国际同行认可则是晋升教授的标准”。不同学校标准虽有不同,但大都没有明文规定 SCI 文章的数量和发表杂志的档次。美国贝勒大学教育学院的学术评价标准则被具体化为如下一些内容^[9]: (1) 问题或目标。研究所要回答的问题或达成的目标是什么,与本领域近期的重大进展如何关联,通过问题或目标确定该研究的类型是什么。(2) 方法或过程。该研究的方法或研究技术路线是否合适,在研究过程中对这些方法或技术的应用是否体现了研究者对该领域知识的认知和掌握。研究者选取了怎样的研究材料和资源来适应研究方法和技术路线,这些研究方法和技术是否严谨和可重复。(3) 成果或绩效。研究成果的呈现方式,研究者如何组织其研究结果,这种组织方式是否有效地向读者传递了该研究的成果和意义。(4) 学者圈内评价的重要性和影响。该领域内的学者如何评价该研究成果,此成果是否影响了领域内的学者对相关知

识的认识,该成果对领域发展的系统性贡献是什么。(5) 同行评价的特征。该研究学术评价的同行是如何产生的,他们是否具有代表性和中立性,这些同行在该领域的学术地位如何,是否适合本次评价。评价的评语和结果是怎样的,有没有与本研究涉及的领域无关的内容,评语是否体现出评价者的专业素质。

实际上,学术评价在欧美国家也是一个颇有争议的话题。他们并非所谓的“过来人”,同样存在很多问题,学术界也一直在争论、改进。例如申请基金时,发表文章多,杂志档次高的申请人明显更容易获得资助。如美国一些大学在考虑晋升时,首先通过获得基金和发表文章情况进行初选,然后组织同行评议。所以其评价机制仍存在过度依赖形式指标的情况。有鉴于此,2012 年美国生物学年会时,与会科学家共同发表了“学术评价(或科研评价)的旧金山宣言, San Francisco declaration on research assessment (DORA)”。宣言主要目的就是提高学术评价的质量,提倡学术评价以研究的内容为依据,减少以 SCI 文章和发表杂志的 JIF 评价学术的情况等。宣言对于学术评价的利益攸关方包括基金提供机构、学术机构、杂志发行方、学术评价机构以及学者的行为都给出了建议(参见 <http://am.ascb.org/dora/>)。旧金山宣言一经发表就受到了广泛的关注,包括很多科学家和著名的杂志如 *Science* 等都对其予以好评,并签署了宣言。当然旧金山宣言并不是金科玉律,比如 *Nature* 杂志就对其中的一些具体的建议有不同的看法,未加入署名行列。

在英国,有专门的官方机构“research assessment exercise (RAE)”,主要负责对研究机构如大学、研究所等进行学术评价(<http://www.rae.ac.uk>)。RAE 从 1986 年开始,到现在已经是 RAE2008。RAE2008 把研究分成 5 个档次,依次是国际领先,国际先进,国际水准,国内水准,无法评价。RAE 主要从研究产出(70%权重)、研究环境(20%权重)、尊敬指数(10%权重)3 个方面进行评价。研究产出都要求已经对公众公开,包括:授权专利、未授权已公开的专利申请、软件、网络出版物、数据库、专著、演出、艺术作品、设计、人工制品、展览、会议的讲演和展板以及各种印刷出版物如各种学术期刊上发表的文章等。RAE 等学术评价机构的运作客观上促进了英国科学研究的良好发展。然而作为最早使用同行评议的国家,在英国同行评议仍然不是一个轻松的话题。由于申请项目的激烈竞争,使同行评议一直处在“紧张气氛”

之中而受到一些科学家的严厉批评. 为此前国务大臣霍恩组织了一个高层次的调查组, 对英国五个科学研究理事会的同行评议进行了全面研究. 调查结果表明“同行评议是一种不可取代的评价基础研究的方法. 当然这种方法不是十全十美的, 尚需引进其他评议方法的优点进行改进”^[16].

为了有效地进行学术评价, 法国于2006年成立了科研与高等教育评价署. 它是一个独立行政机构, 负责组织对由政府资助的所有科研计划和项目进行评价, 并向社会公布评审结果. 具体做法是: 建立独立、公开、透明的评价程序, 按照通用的国际标准, 在完全独立于主管部门和评审对象的前提下开展同行评议(领域专家评价项目)、质量评价(国际知名专家标准方法)、“透明评价”(根据公开标准)和“真实评价”(以无可争辩的事实为依据).

4 我国学术评价的现状

2009年全国科学技术协会年会开幕式上, 韩启德院士对中国学术评价的现状进行了总结: “近年来, 我国科技评价体系不完善的问题开始凸显, 中国科学技术协会(简称科协)调研显示部分科技工作者对现行评价体系的不信任感和挫折感相当强烈”. 科协调研发现在政府科研经费资助和项目管理上, 官员的权力过大, 影响评审的科学性和公正性; 在人才评价和科技成果、资助项目评定上, 缺少严格意义上的同行评议, 把论文发表数量、引用率等量化指标作为唯一标准, 形成“学术工分”式的评价方法; 在科技奖励上, 重政府奖励, 轻学术共同体奖励, 两者缺乏衔接, 一些政府奖励的同行认可程度比较低. 学术评价的这些问题导致了科研资源配置不尽合理, 科技经费使用的效率不高, 而且助长了一些不良的风气, 不利于学术道德的建设, 也不利于创新热情和创造活力的激发. 韩启德院士强调, 这些问题之所以存在, 关键在于“学术共同体没有能够在学术评价中发挥应有的作用, 学术共同体的主动性、责任性、积极性和自身能力都不够.”

当具体分析学术共同体失职的原因时, 不同程度地存在以下一些问题:

4.1 学术评价价值导向失范

科学研究是一项以艰苦探索为基本特征的高度创造性的实践活动. 早在20世纪30年代美国社会心

理学家 Robert Merton 就提出了科学家应该遵循的4项规范(或科学家必须具备的“精神气质”), 包括: 普遍性, 科学的评判标准是事先确定的, 不受非科学的人为因素干扰; 共有性, 科学发现归科学共同体所有; 无私利性, 科学研究的出发点是推动科学发展, 而不是获取个人私利; 有理性的怀疑, 科学家要敢于对科学发现提出质疑^[17]. 这4种精神气质构成了科学共同体的共同价值规范体系, 科学家在不同程度上将这些规范内化为自己的科学良心, 规范自己的科学行为. 科学系统正是通过这4种精神气质构成的价值规范体系来进行社会整合和模式维持, 以实现科学目标. 然而目前国内学术评价简单地将科学研究这种高度创造性的实践活动当作一般的物质生产活动进行管理评价, 将其量化为科研项目数量、科研项目级别、发表论文数量、级别等, 并将量化的结果简单地与职称评定、福利津贴等个人利益直接挂钩^[18]. 这样的学术评价和奖励方式一定程度上导致科研价值导向失范, 间接引导科研人员急功近利, 出现避难就易的倾向, 一定程度上妨碍了科学发展. 应该指出的是科学 Merton 规范只是科学研究及学术评价的一种保障性措施, 是科学家从事科学活动的原则, 遵循它不会受到奖励, 因为社会实践中“并不存在对遵从任一套社会规范提供奖赏的制度性机制”^[19]; 而违背它则会遭到谴责.

4.2 评价主体的行为缺乏规范

科学是科学共同体共同的事业, 学术评价是科学共同体共同义务. 目前, 在世界范围内学术评价很大程度上还是依靠每次具体参与评议的少量同行. 这种情况下, 学术评价的质量很大程度上依赖于同行具体的评价行为. 前面提到评议人的评价质量受到评议人主观和客观因素的制约, 如评议人的知识结构、学术兴趣、道德心理、人际关系、工作责任心等. 多数评议人都是普通的科学家, 因此一些开拓性的论文和大有前途的研究有可能不被理解, 遭到压制. 例如美国海洋研究局、原子能委员会和国家科学基金, 就都以过分思辨的理由拒绝了 Donald Glaser 要求资助“气泡室”的申请. 然而就是在这个气泡室的基础上, Glaser 发明了最重要的高能粒子探测器, 并由此获得了1960年的诺贝尔物理学奖. 同行评议制度还常常被指责为像“朋友圈”. 科学家之间可能会出现利用同行评议互惠互利的情况. 另外, 同行评

议还可能出现所谓学术评价的“马太效应”，即著名科学家容易获得与被评价项目水平无关的好评，这现象对青年科学家尤为不利。统计显示，大部分诺贝尔奖获得者其获奖学术成果是在科学家处于中青年阶段获得的，也就是说中青年科学家是最有创造力的群体。可以说同行评议的“马太效应”对科学的发展会产生阻力。因此，选择恰当的评议人，使评议主体的主观和客观条件最适合当下的学术评价，并在学术评价过程中和之后对评议人的行为进行规范，是避免上述问题，提高评议质量的关键。

我国的学术评价包括同行评议对于评价的主体即评议人还缺乏一些规范，如评议人的选择和评议人的行为等。由于评议人的选择不当，或评议人的一些不恰当的评价或与评议本身没有直接关系的行为就可能导致了评议的质量受到干扰，使被评议人或评议项目即评议的客体受到不公正的待遇。久而久之，学术评价的权威性和公正性就会受到质疑。

4.3 评价标准不合理

学术评价最根本的价值在于科学价值。在改革开放初期，受限于我国学术共同体自身的主观和客观条件，在我国逐渐形成了以SCI论文数量和发表期刊影响因子为量化依据的学术评价标准。目前我国每年发表的SCI论文数量已经仅次于美国，现在如果仍以这种形式化的评价方式为主，会造成科研人员过于注重SCI论文数量和杂志档次，忽视论文背后所代表的真正科研水平和科学价值，影响我国科研水平的进一步提升。另外，科学研究一般包括基础研究和应用研究。研究类型不同，其评价指标、评价方式也应该有所差异。一般来说应用性的科学研究出成果往往比基础研究要慢。以转化医学为例，一个成功的转化研究一般需要一个团队耗时10~15年^[20]，而一篇高水平的科学文章，如*Cell*, *Nature*, *Science*这些杂志的文章一般平均耗费一个科研人员3~5年左右的时间。过于注重SCI论文的评价方式客观上造成了对从事应用研究的科研人员的不公正，打击了科研人员从事应用研究的积极性，使得大批研究人员只做基础研究，发表论文，而不愿从事成果的转化研究。实际上，这也一定程度上违背了科学研究的最终目的，即促进人类社会实践活动，提升人类生活质量。另外，还有一个情况就是把学术评价应该采用的科学标准和一些技术职业评定标准混为一谈。比如医

学研究领域除了专职研究人员外还有大量临床医生做兼职研究人员。很多医学院采用基础和临床学术评价两条线，主要表现在临床医生作学术职称评定时，普遍采用低于基础研究人员的评定标准，很多时候甚至采用或部分采用医师职称评定标准作为学术职称的评定标准。结果同一学术机构里拥有同等学术职称人员的学术水平差别很大，这实际影响了学术职称和学术评价的严肃性。

4.4 科研评价机制不健全

随着我国科研水平的迅速提升，现有的学术评价机制与科研形势不相适应的问题日渐凸显。主要表现在：学术评价管理单位的权利过大，能够较大程度上影响评价的结果，从而降低了评议人参与的积极性，降低了学术评价的学术含量；学术评价评议人的管理不规范，包括评议人的选择、评议人评议行为的约束、评议人非评议行为的约束等，影响了学术评价的公正性。学术评价的评审过程透明度不够，影响了学术评价的公信力；评价结果的反馈机制不完善，评审专家信誉制度尚未建立，减低了学术评价的自我纠错能力。另一方面，现代科学的发展日新月异，而我国科学技术发展由于是后发的追赶状态，所以我们发展的相对速度更快。这就对学术评价管理人自身的素质提出了更高的要求，不仅要懂管理还要懂科学，对自己管理领域的最新进展和其意义都要有深刻的认识。有些评议人受自己知识结构、学术水平、科学兴趣的限制很难对所有的学术评价都做出高质量的判断。这时候就需要学术评价管理人对领域内评议人的学术水平、知识结构、学科兴趣有清楚的了解，对评议人评价的价值标准的科学含义有较为清楚的认识，能在适当的时候对评议人和评价过程进行恰当的调整。这实际对学术评价管理人提出了相当高科学素质的要求。

5 对我国学术评价的一些建议

5.1 评价方式上基于内容的评价和形式的评价相结合

我国的学术评价要向基于内容的同行评议转变，但不能排斥基于形式评价方。学术评价采用同行评议时需要考虑的是该领域有没有足够的高水平科学家，他们的知识结构、学术兴趣、道德心理、人际关

系、工作责任心等能不能胜任。一般来说国家级的学术奖励都能够找到足够的高水平科学家进行评议。但在没有足够的高水平科学家参与评价时，引入基于形式评价方式是保证评议质量的重要手段。基于形式的学术评价很多时候是学术标准的量化，它是通过对各种评价标准设置不同权重而实现的。指标体系的设计应注重成果引用等客观量化评价、经济产出、社会影响等指标的设置。另外，适当的引入基于形式评价方式是防止学术评价中学术腐败，少数人把持学术界话语权和学术评价马太效应的有效方法。因此，我们主张在回归基于内容的学术评价的方法的同时，需要考虑根据当下学术评价的具体条件适当的引入基于形式评价的方法。实际上，就基础研究来说，文章发表在哪里，文章发表后的引用情况，特别是对文章中心议题支持的引用，一定程度上在更大的范围里反应了学术界对这篇文章科学意义的认可。在引入基于形式评价时，应该更多的关注对文章中心议题支持的引用，适当淡化文章发表的杂志的 IF。

5.2 评议人选择上小同行和大同行相结合

现代科学的发展，在多学科综合趋势增强的同时，专业研究领域的划分越来越细，“隔行如隔山”的知识障碍愈益明显，“无所不知”越来越难。研究领域越相近，研究方向越一致的同行就越能够充分理解评议客体的科学价值。因此，我们在进行同行评议时首选小同行，比如组织二、三级学科或专门研究领域的小同行专家评审。特别是进行通讯评审时，可在同一学科或跨学科对最窄的研究领域选择评审专家，适应了科学研究既综合又分化的发展趋势。另外，“SCI之父”加菲尔德提出如果深入挖掘 SCI 系统的功能可以找到更好的评估方法来纠正现有的谬误。比如将科研人员的研究成果放在他所在的研究领域中进行评估，即先在数据库中寻找其小同行群体，然后让这一群体来衡量此人的贡献^[21]。这种方法的好处是能够根据科学家发表文章的领域和数量来确定其真正专注的领域和活跃程度。当然这需要大量的文献计量和分析，建立完备的不断更新的数据库。然而小同行的认识也会有局限，也会犯错误。一是过度拔高评议对象的科学价值。每个科学家都专注于某个或某些研究方向，并且倾向于认为自己做的是重要的研究。因此，某个小领域的研究群体可能会倾向于

认为自己的领域的任何进展都具有重大的科学价值，从而可能过度拔高评议对象的科学价值。因此，就需要适当的让一些大领域里领先的大同行参与评议，使评议能够在更广大的范围里符合其客观现状。二是故意贬低评议对象的科学价值。存在竞争关系的小同行之间有可能出现贬低评议对象的科学价值的情况，尤其是竞争激烈的领域。这种情况对于没有根基正在成长的年轻科学家尤为不利。这时如果有一些大同行，由于没有直接的利害关系和学识思想的碰撞，其评价往往会较少受非学术因素的左右。

5.3 被评议人举证和评议人评议相结合

科学是科学共同体共同的事业，那么维护科学正确的价值导向则是科学共同体共同的义务。就学术评价来说每个科学人都只有义务没有权利。然而每个科学家也是自然人，只承担义务不能用之谋利则不可能要求其投入太多的时间与精力。那么如何通过减轻评议人的工作量来保证评议质量就有现实的需求。前面提到小同行评议是因为一般来说小同行是最能理解评议客体科学意义的一群人。实际上，被评议人也是最能理解评议客体科学意义人。因此，让被评议人举证其研究项目的科学意义是一种高效的方法。比如最近，美国国立癌症中心的主任 Harold Varmus 就提出要求研究人员在申请课题时在其简历中描述其最重要的研究成果和工作而不是仅仅罗列已发表的文章。当然，被评议人举证并不是自吹自擂，而是要在客观描述自己科研成果的同时，使用已有的证据或客观指标来为评议人评估其科学意义提供线索，减少评议人调研的时间，帮助评议人在较短时间内全面了解评议客体的相关信息。另一方面，评议人可以采用被评议人提供的信息，也可以自己收集相关信息为最终客观的评议提供依据。

5.4 评价标准中科学意义和社会经济价值并重

学术评价的同行评议主要是要求领域内的专家对评议客体的科学意义做出评估和预测。好的研究一定具有普遍意义，会被其他大量研究重复和借鉴，包括学术思想，研究模式和具体结果等。然而即使是最优秀的科学家也不能完全准确预测一项研究的科学价值。所有的研究都有待时间的检验。对诺贝尔奖近几十年授奖情况的统计表明，研究成果从发表到获奖的平均时间间隔，物理学是 13.1 年，化学为 14.3

年,生物学和医学是14.2年^[22]。就一篇论文来说,其98.5%引用发生在文献发表年份的2年之后,其中数学和物理科学领域对发表两年内的文献的引用率仅1%~3%,而生物科学领域则为5%~8%^[13]。因此研究成果被认可往往需要一定时间。在考虑研究的科学意义,特别是进行国家级的学术评价时,最好不要急于对一项刚发表的工作做出评价。

科学以其研究成果提升社会实践为目的。任何科学研究都需要真正的施惠人类社会才是其最终价值的体现,所以社会经济价值实际是学术评价的终极标准。基础研究在没有真正转化前,往往难以准确评估其社会经济价值,应用研究或转化研究的社会经济价值则相对容易的判断。但是成功的应用研究或转化研究往往耗时数十年,如何在研究处于中间状态时正确的评价其价值,以便这些研究能得到持续的关注和支持,仍然是需要思考的问题。笔者认为可以根据每类研究的特点及其客观规律,设定一些里程碑式的节点,每个节点设置一些关键的成功标准进行评估。显然,越靠后的节点其意义越大。

5.5 完善评价的管理和反馈机制

学术评价的管理机制的建立和完善是制约我国学术评价向更高水平发展的一个短板。学术评价的管理机制包括两个方面:(1)科学共同体在学术评价中的主体性地位的完善和评议人行为的约束;(2)学术评价机构或曰中介结构在学术评价中角色的明确和行为约束。

学术评价中体现的科学价值应该是整个科学共同体的价值取向,这就需要科学共同体在学术评价中成为真正的裁判员。在建全学术评价的管理机制时,需要明确评议的同行群体作为学术评价最终裁决人的地位,学术评价不要受到行政等非科学价值判断的干扰。学术共同体要参与制定学术评价的机制,包括学术评价的范式、标准、流程等所有细节,并对这些规范的制订享有最终决定权。同时需要明确参与评议同行的义务,包括保守评议客体信息的义务,未经授权永久保密其作为某项学术评价同行评议人身份以维持同行评议单盲原则的义务,利害冲突自动回避的义务等。

另一方面,学术评价机构应该是学术评价的发起人,根据国民经济的需要和国家的科技政策发起某项学术评价,汇同科学共同体共同完成该项学术

评价的同行评议人信息数据库建设,同行遴选标准、评议人行为标准、学术评价范式、标准、流程等的制度建设。实际上,学术评价的制度建设一直是我国的弱项,很多学术评价只有大概的原则,没有操作的细节,这就给了包括评议人和行政人员很多操作的空间,使学术评价受到非科学因素的干扰。学术评价机构还应该是学识评价的管理人,负责推动和监督学术评价按既定的制度执行,同时还要推动学识评价的透明化,使学术评价的过程和每个阶段性决议及其依据对科学共同体公开。即在不涉密的情况下,每个科学人可以通过公开的渠道方便地获取正在进行的学术评价的相关信息,包括匿名评议人的具体评语等,使得没有参与评议的同行和被评议人也可以评估评议的客观性,并对其中明显的错误实时提出异议。学术评价机构还应该是学术评价的监督人,要对评议过程中的违规行为进行干预,并负责将没有参与评议的同行提出的意见及时传达给所有评议人。但是很明显学术评价机构一定不是学术评价的执行人。

由于评议人在其他评议活动中也可能成为被评议对象,因此学术评价的反馈机制对于保证学术评价的质量,防止学术腐败就极为重要。建立学术评价的反馈机制主要是在学术评价完成后对每次学术评价中出现的问题进行总结,并在评议的制度建设中得以体现。依据评议人在学术评价中承担的义务,学术评价的制度和没有参与评议的同行对评议的意见对每次参与评议的评议人行为和质量进行评估,对做得不好的评议人要提出意见,必要时要予以撤换。依据学术评价的制度对学术评价机构内部具体参与某项评议的管理人员行为进行评估,使评议管理人员即付起的相应的职责,又没有超过应有的限度。当然,科学管理人员能否充分地负起其职责,不仅取决于其责任心,更依赖于其科学素质。因此,学术评价机构应该把专职管理人员的科学素质培养和考核作为重要的人事工作,同时适当增加兼职管理人员,保证管理人员在科学共同体内的合理流动,形成内行管内行的局面。

6 医学科学学术评价建设的一些想法

医学科学是获得人类自身的相关知识,用以促进健康的科学。其既包含获取健康相关知识的基础医学研究,也包括促进健康为目的的应用医学研究,

如临床医学、护理医学、循证医学等研究,同时还有如何把基础医学研究成果转化为实际应用的转化医学.因此,医学研究学术评价的标准应包含理论创新的价值判断,也应突显提升疾病预防和治疗水平带来的社会、经济价值的考量.在医疗界经常问到的问题是,没有发表论文或著作的医师能不能是好医生,回答当然是肯定的.但是,很多时候特别是医学院附属教学医院的医生不仅要取得医师职称,还要求获得学术职称,这种情况就应该要求有学术成果.很多时候高级执业医师和学术职称是挂钩的,比如哈佛医学院主要附属医院的科主任医师一般也是医学院的教授.

另一方面,在考虑医学科学学术评价时还要考虑医疗行业本身的特点给医学研究带来的特殊性.这体现在医学研究面对的是人类本身,不管是医学知识上的扩展,操作技术的进步,新型疗法的推广等都是为了提升人类的健康水平.除了专门从事医学研究的科研人员外,还有数量庞大的临床医生和卫生管理人员队伍,如何提升这些人群的科研积极性和科研水平是医学科学学术评价建立所必须考虑的问题.另外,医学研究包括基础研究、预防研究、临床研究、转化研究等,从事医学研究的人可以专注于其中任何一种研究,不管他本人是基础医学工作者还是临床医学工作者.因此,医学科学学术评价的标准应该针对不同的研究类型有所区分.为了保证学术评价的严肃性和统一性,学术评价标准不应该由于学术活动从事者的专职职业不同就有所区别,不管是临床医师或是基础医学工作者只要他们申报的研究类型一致,比如都是基础医学研究,就应该采用相同的标准.基础医学研究以拓展人类对健康的知识为任务,则评价应基于科学内容,在某些场合可适当考虑文章的被引用情况和文章发表的杂志水平等客观评价指标.临床医学研究成果有些发表了很好的论文,有些发表的论文水平可能一般,有些甚至没有正式论文发表,但对临床产生的影响却可能和论文发表水平没有直接关联.因此临床医学的学术评价标准应该更关注该研究成果在多大程度上获得医疗界同业的认可,比如某种新的疗法、新的术式,某种疗法和术式的改进或某种治疗方案应用范围或适应症的扩展等,是否被同业者广泛认同和采用,在多大程度上提高了医疗质量和产生了多大的社会经济价值.另一方面,医疗质量提升带来的社会经济价值

也可以在推广某种新的疗法、新的术式中产生,尽管这并非推广者的原创成果.这里我们认为是否被同业者广泛认同和采用是最关键、最客观的标准,要避免自卖自夸式标准.

医学研究中的转化医学研究,其目标是将基础医学研究成果转换为临床可用的技术规范、药物、设备等^[20,23].一般而言一项成功的转化耗时约10~15年,一项基础研究或临床研究一般耗时3~5年.如果一定要等到一项转化研究完成后再进行学术评价,这就对从事转化研究的科研人员不公平.因此建议设定转化研究或应用研究的里程碑式的学术评价标准,比如(1)转化医学研究核心专利的获得,(2)申报临床试用获得批准,(3)二期临床试用完成,(4)三期临床试用完成等.需要指明的是转化医学研究核心专利是指能都对以后将要上市的产品提供直接知识产权保护专利.靶标专利,分子标志物的分子专利都不是核心专利^[23].

可以看出,基础医学研究学术评价着重于评价对象的科学性,采用评议同行的主观标准,而临床医学和转化医学研究的学术评价看重学术成果外在表现方式,以评价对象的社会和经济价值等客观标准为主.那么对应于不同的评审标准,在进行学术评价时,可能需要设置不同的流程,不同的评审专家.这些在医学学术评价的制度建设时都应该加以考虑.比如,与专门基础医学研究的科研人员相比,大多是临床从业人员从事基础医学研究没有太多优势,但是从事临床或转化研究就有较大的优势.因此,医学学术评价的奖励机制,如课题申请、学术荣誉等可适当向临床或转化研究倾斜,以鼓励临床从业人员参与这些研究.

7 结语

科学是所有科学人的科学,是科学共同体共同的事业,事关全人类的福祉.纵观人类历史,科学技术的发展始终是人类社会进步的根本推动力.随着生产力不断提高,将会有越来越多人进入科学技术研究的领域,科学人的群体也将越来越大.如何实现科学共同体的自我管理,保证科学系统高效良性运行,不仅是科学界的问题,也将是整个社会的问题.学术评价是科学共同体共同科学价值观的体现,是科学共同体自我管理建设中的灯塔和风向标.本文想要探讨的正是我国学术评价方法的改进和相应的

制度建设. 建立学术评价制度是为了保证学术评价的科学性, 防止非科学因素的掺入. 我国在很多社会实践中存在重标准、轻程序的现象. 在讨论学术评价改革时也可能存在这一问题. 在把评议标准改为基于科学内容的主观标准而非 SCI 和 JIF 等客观标准时, 如果没有周密的程序设计以引导和制约评议主体的评价活动, 就可能导致主观随意性的泛滥而减低学

术评价的科学性. 当然我们也始终坚信科学人作为全社会最理性、最有献身精神的一群人, 一定能够担当起应有的角色. 正如 June Goodfield 曾言: “在所有职业中, 科学最富于批判性. 有专职的音乐批评家、美术批评家、诗歌和散文批评家, 但就是没有专职的科学批评家, 因为科学家自己就可以胜任这一角色.”

参考文献

- 1 中国科学技术信息研究所. 2012 中国科技论文统计结果
- 2 我国发明专利授权大幅增长——全国政协委员、国家知识产权局局长田力普等谈专利. 中国科技网, 2013 年 3 月 12 日
- 3 张保生. 学术评价的性质和作用. 学术研究, 2006, 2: 10-15
- 4 科技服务业结构的重大变化及发展重点. 人民网, 2012 年 12 月 17 日
- 5 培根. 学术的进展. 刘运同, 译. 上海: 上海人民出版社, 2007
- 6 俞吾金. 也谈学术规范、学术民主与学术自由. 学术界, 2002, 3
- 7 武小琳, 钱文霖. 保证学术期刊审定稿科学性的若干理论思考. 编辑学报, 2004, 16: 4-6
- 8 Shulman L S. Taking learning seriously. Change, 1999, 31: 10-17
- 9 刘剑虹, 吕杰. 学术评价: 美国大学的一种理论模式——以贝勒大学为例. 宁波大学学报: 教育科学版, 2002, 5: 1-4
- 10 Garfield E. Citation indexes for science: A new dimension in documentation through association of ideas. Science, 1955, 122: 108-111
- 11 Garfield E. Citation analysis as a tool in journal evaluation. Science, 1972, 178: 471-479
- 12 Seglen P O. The skewness of science. J Amer Soc Inform Sci, 1992, 43: 628-638
- 13 Krell F T. Why impact factors don't work for taxonomy. Nature, 2002, 415: 957
- 14 Garfield E. The impact factor and using it correctly. Der Unfallchirurg, 1998, 48: 413-414
- 15 约翰·罗尔斯. 正义论. 何怀宏, 何包刚, 廖申白, 译. 北京: 中国社会科学出版社, 1988
- 16 吴述尧. 同行评议的对比研究. 中国科学基金, 1993, 2: 136
- 17 Merton R K. The Sociology of Science: Theoretical and Empirical Investigation. Chicago: University of Chicago Press, 1973
- 18 汪静. 高校科研评价体系的现状及对策. 中国高校科技, 2012, 290: 12-13
- 19 Mulkay M. Sociology of Science: A Sociological Pilgrimage. Indiana: Indiana University Press, 1991
- 20 Contopoulos-Ioannidis D G, Alexiou G A, Gouvias T C, et al. Medicine. Life cycle of translational research for medical interventions. Science, 2008, 321: 1298-1299
- 21 付雁南. “SCI 之父”加菲尔德: SCI 畸形地位是必经阶段. 中国青年报, 2009 年 9 月 16 日
- 22 张彦. 论科学评价的社会学机制. 南京大学学报(哲学社会科学版), 1995, 4: 60-66
- 23 董尔丹, 胡海, 洪微. 浅析转化医学与医学实践. 科学通报, 2013, 58: 53-62

To improve the research assessment process in China

DONG ErDan, HU Hai & ZHANG Jun

Department of Health Sciences, National Natural Science Foundation of China, Beijing 100085, China

Research assessments make value judgments on the results of academic activities. In China, the existing research assessment models have become increasingly unsuited to the rapid development of science and technology. Here we discuss the essence and basic principles of research assessment. We demonstrate the characteristics, advantages and disadvantages of two popular models for research assessment: one based on the scientific content and the other based on the form of academic achievement. We then review the current situation and problems of research assessment in Western countries, as well as the current problems in China. We propose the principles and corresponding institutional improvements that should follow upon changing research assessment from a form-based to content-based model. We also discuss the changes and improvements that should be considered in medical research assessment.

research assessment, peer review, scientific value, science community

doi: 10.1360/972013-1245