中国科技资源导刊 ISSN 1674-1544
2012年7月 第44卷第4期 91-97
CHINA SCIENCE & TECHNOLOGY RESOURCES REVIEW
ISSN 1674-1544 Vol.44 No.4 91-97, Jul. 2012

基础医学研究成果的综合评价

杨 颖[1] 黄亚明[1] 何钦成[2]
（1. 中国医科大学图书馆，辽宁沈阳 110001；2. 中国医科大学，辽宁沈阳 110001）

摘 要：构建以科学计量学指标为基础的基础医学科研成果的评价指标体系,并用Logistic回归与判别分析模拟评价,比较两种方法在基础医学研究成果评价中的评价效果。结果表明，在预测准确率方面，Logistic回归比判别分析稍高，两者筛选出了经验认为重要的变量，较全面反映科研成果水平的各个方面，初拟评价指标体系合理，可用于评价基础医学科研成果，即为成果评价的初步预评提供参考。

关键词：科学计量学；科研成果评价；Logistic回归；判别分析；基础医学
中图分类号：R181.3 **文献标志码：**A DOI：10.3772/j.issn.1674-1544.2012.04.019

Multiindex-oriented Evaluation of Scientific Research Achievement for Preclinical Medicine

Yang Ying[1], Huang Yaming[1], He Qincheng[2]
(1. Library of China Medical University, Shenyang 110001; 2. China Medical University, Shenyang 110001)

Abstract: The purpose of this study was to construct an evaluation system of preclinical medical research achievement which comprised of indicators based on scientometrics and to validate effects of the by simulating the evaluation of peer review through logistic regression and distinguish Analysis. Resulting indication: the predicted percentage of the two methods was that the logistic regression is more slightly high than the distinguish Analysis. Important variables which is identified from obtained experience in both, reflect roundly various research harvests, is reasonable for decided evaluation system, to use evaluation of preclinical medical research achievement.

Keywords: scientometrics, evaluation of scientific research achievement, logistic regression discriminant analysis, Preclinical Medicine

1 引言

在当今科技迅猛发展的时代，科技评价的重要性得到了普遍的重视。明确科学研究成果评价的价值标准，是科学地评价高等学校、科研单位及其教师和科研工作者教学与科研水平的重要依据，使科研人员客观地了解自身的水平和学术影响，促进科学的发展和社会的进步[1]。根据科学技术工作的性质，通常把科研成果分成基础理论研究成果、应用技术研究成果和情报调查研究成果三大类[2]。而对于科研成果质量的评估，长期以来各国最基本、最常用的是同行评议(peer review)的方法，但是由于同行评议在本质上是个人意见的表述，人为因素不可避免。随着科学计量学的发展，人们应用科学计量学方法客观地定量地对学术论文水平、期刊质量进行评价，进而对科研成果水平进行评价。这种评价方法克服了同行评议的弊端，也从一个侧面判断某个国家科技发展水平、动向及发展潜力，为决策与规划提供参考依据[3]。在对科研成果的评价进行定量分析时,《科学引文索引》(Science Citation Index，简称SCI)是目前国际上广泛承认的一种科研成果评价体系。科研机构被SCI收录和引用的论

第一作者简介：杨颖(1980-)，女，中国医科大学图书馆助理研究员，研究方向：文本挖掘。
收稿日期：2011年1月28日。

文总量，反映整个机构的科研尤其是基础研究的水平；个人的论文被SCI收录的数量及被引用次数反映个人的研究能力与学术水平[4]。但是，科研项目发表的成果不仅反映在论文发表、论文被引、被检索工具收录等方面，也体现在专利、参编专著等方面。因此，不能简单地以这种期刊源的统计结果直接对科研成果进行评价。在以往实践中发现，采用单一的方法和指标评价科研成果，往往带来片面性和误差。综合采用多种方法和多项指标评价科研成果已成为一种趋势。将以定性为主的同行评议方法与定量为主的科学计量学方法有机结合起来进行评价的方法，已被用于评价科研活动效果、决定基金资助、职称评定和科研人员的绩效评价等领域，并取得了良好效果。

在基础医学科学研究中，通常对计量指标进行定量分析，尤其是数理统计方法的应用，如回归分析、判别分析等，能够使定量研究获得更为科学、客观、可靠的结果。本研究拟应用科学计量学指标建立基础医学研究成果的评价系统，应用Logistic回归和判别分析对基础医学研究成果进行评价分析，验证以科学计量学指标为基础的评价指标体系的效果。

2　评价指标

本研究以国家某部2003年基础研究成果奖全国推荐评奖项目中基础医学类项目共18项为样本。

2.1　指标的拟定和选择

初步拟定9个指标：①承担课题，即基金资助情况分为国家重点项目、杰出青年基金项目、自然科学基金重点项目、自然科学基金一般项目、部委项目、国际合作项目、省级项目、国内合作项目、市厅级项目9个亚分类；②项目执行时间（年）；③论文发表；④参编专著（3个亚分类即分别统计项目完成人担任主编、副主编和编委的人次）；⑤获取专利（统计发明专利、实用新型专利、外观设计专利和申请专利的个数）；⑥会议交流情况（4个亚分类即分别统计国际特邀报告与主报告、国际一般会议交流、国内特邀报告与主报告、国内一般会议交流以及国内其他会议交流的篇次）；⑦论文被期刊论文引用；⑧论文被专著综述引用；⑨论文被检索工具收录（两个亚分类即论文被国际性检索工具如SCI、IM、EM、BA、CA等收录和被国内检索工具如CSCD、CNKI收录的篇次）。

本研究中选用这9个指标，是因为：科技论文的数量和质量，一定程度上反映了科学研究成果的水平，尤其是基础研究，这一点目前在科研评估领域已基本形成共识。由于科技期刊质量主要是由所发表论文的质量决定的，反映期刊水平的期刊影响因子在某种程度上可以间接反映出论文质量。但是，期刊影响因子本身存在一些不足，影响着评价科研成果的可靠性。首先期刊的影响因子不等于论文的影响因子，其次学科的大小和特点影响到影响因子的大小。Seglen在前人的基础上指出，某些领域比如临床科学，会引用基础医学的文章，但反过来则不然。因此推断基础医学的平均影响因子可能较高。直接用影响因子来比较不同的学科即使是相近的学科也是不可靠的[5]。为避免影响因子存在的以上问题，本研究查找出该论文在不同领域中的相关文献，计算该论文发表的期刊的影响因子在与其相关文献所在期刊的影响因子中所占的百分位数，为论文发表情况赋分。这样，就知道该论文在不同领域不同学科所占的地位，从而在一定程度上消除了影响因子在学科差异方面的局限性。

成果的表现形式除发表论文外，还有参加会议交流以及参编专著情况。医学学术会议论文是医学文献的重要组成部分，也是重要的信息源。其信息量大、内容新颖、集中、专一，仅次于期刊，并能反映出当前教学、科研和临床发展的动态和水平。另外，在科学成果的产出中，专利也是一个非常重要的方面。申请专利就意味着占有这个专利技术的市场，如果急于发表论文而没有及时申请专利，就等于公开了技术秘密，可能将研究成果拱手相送他人[6]。所以获取专利也是成果的重要组成部分，应作为科研成果的评价指标。

一个项目成果的论著被同行引用的情况往往代表该成果在学术界造成的影响和受重视的程度，所以在评价科研成果时，论著被引也不失为一个重要指标。论文被引用次数反映了该论文的使用价值以及该论文在本学科研究的活跃程度，从而反映了该论文的学术水平，是论文评价的重要指标。此外，论著被专著和综述引用也是学术影响的一个方面。论文被检索工具收录一方面反映了论文的学术水平，另一方面也反映了论文可能造成的学术影响。

承担课题情况反映与科研成果直接相关的研究工作的起始状况，即该项目立项时的先进性和新颖

性，其拟解决的科学问题对社会、经济、科技发展的重要性。项目的执行时间则反映成果的工作量，也可以作为评价科研成果的指标。

2.2　指标的筛选

本研究中得出的4个结果中均纳入论文被引，表明论文被引是评价基础研究成果最重要的指标。本研究在统计论文被引时，排除了自引。这个指标在几种方法中均入选，是因为论文被引情况反映该成果在学术界的活跃程度、受重视程度以及所造成的影响。3个结果中纳入论文被引、被专著综述引用和项目执行时间。项目执行时间反映工作量，在某种程度上也反映出项目难度。被专著综述收录反映出论文受重视的程度以及论文的水平。以上指标被多种方法纳入评价体系，这与目前评价工作的实践相符合。两个结果纳入了论文得分、课题、专利。论文得分，说明本研究对此指标的处理在一定程度上反映了论文的数量和质量两方面特征，但Logistic回归结果没纳入此变量，可能是Logistic回归本身易受共线性影响所致，而论文被引、论文发表、被检索工具收录3个指标间必然存在共线性关系。承担课题情况体现项目的新颖性，专利体现了它在基础医学科研成果评价中的重要性，这些指标的纳入符合当前成果评估的实践经验。

2.3　成果发表情况的指标

以被评价研究成果提供的论文的关键词为检索策略，在CBM和Pubmed数据库中分别检索出每篇论文从1995年至2003年的中、英文相关文献。通过CSCD和JCR统计这些论文所在期刊的影响因子。若期刊未被这两个数据库收录，则为其赋IF值0.0005。编程计算项目发表论文所在期刊的影响因子在相关文献刊载期刊的影响因子中的百分位数[7]，将每个项目的所有发表论文的百分位次相加求和，即得项目发表论文得分。

2.4　分级指标

用孟凯韬提出的赋权法[8]分别为论文被检索工具收录、获取专利、会议交流、课题、参编专著5个指标的不同亚分类赋权重，再用权重乘以每个亚分类的个数，求和为此项目该指标得分。

另外，项目执行时间、论文被期刊论文引用和被专著综述引用取实测值。

为使各个项目对应的指标值分布相对集中，故对各个指标值进行0−1处理，即对于每一项指标，用原值除以各项目这一指标值中的最大值。

3　Logistic回归分析

本文选用SPSS11.5/WIN中Logistic回归分析的前向逐步进入法和STATISTICA软件中判别分析法的后向逐步判别分析法[9]，将上述9个指标作为自变量，获奖等级作为应变量（Y）。

按照两种评议思维方式进行模拟，一种是评价拟授奖与不授奖（简称方法一），不授奖Y值记为0，拟授奖Y值记为1。另一种评价方法（简称方法二）是评价一等奖与非一等奖，一等奖Y值记为1，非一等奖Y值记为0。

为使各个项目对应的指标值分布相对集中，故对各个指标值进行0−1处理，即对于每一项指标，用原值除以各项目这一指标值中的最大值，所得结果见表1。

应用SPSS软件中Binary Logistic分析方法，采取前向逐步回归进行了分析，分别得出两种评价方式的Logistic回归分类预测结果（概率分界点置为0.5）和变量纳入结果。

3.1　方法一前向逐步回归结果

（1）分类结果：前向逐步回归总共进行了六步，第五步的准确预测率最高，为94.44444%，第六步的准确预测率降为83.33333%（表2）。

（2）变量结果：第一步进入方程的变量：论文被引；第二步纳入的变量：专利、论文被引；第三步纳入的变量：专利、论文被引和被专著综述引用；第四步纳入的变量有：专利、课题、论文被引和被专著综述引用；第五步纳入的变量：专利、课题、项目执行时间、论文被引和被专著综述引用；第六步纳入的变量：专利、课题、项目执行时间、论文被引（表3）。

3.2　方法二前向逐步回归结果

（1）分类结果：前向逐步回归只进行了一步，预测准确率为100%（表4）。

（2）变量纳入结果：进入方程的变量为论文被引（表5）。

4　Logistic判别分析

应用STATISTICA FOR WIN软件中DISCRIMINANT ANALYSIS分析方法，前向逐步回归进行了分析，进入模型和排除的F值标准分别为1和0.5，得出两种评价方式的判别分析的分类预测结果和变量纳入结果。

表1　分析前各个项目各指标分值

项目编码	专利	会议交流	课题	参编专著	被检索工具收录	项目执行时间	论文被引	被专著综述引用	论文得分	Y
310001	0	1	0.1071	0	0.4719	0.1667	0.0000	0	0.5522	0
310002	0	0	0.1071	0	0	0.2083	0.0000	0	0.0768	0
310003	0	0.145	0.2984	0	0.1049	0.3333	1.0000	0	0.2888	1
310004	0	0	0.4284	0	0.2098	0.75	0.0341	0	0.5071	0
310005	0	0.022	0	0	0.0262	0.9167	0.0000	0.3333	0.1329	0
310006	0.375	0.162	0.5223	0	0.236	1	0.0000	0	0.9107	0
310007	0	0	0.2115	0	0.0988	0.2083	0.0000	0	0.0573	0
310008	0	0.159	0.8639	0.1052	0.4195	0.375	0.0000	0	0.6974	1
310009	0	0	0.8066	0	0.1573	0.375	0.0000	0	0.452	0
310010	0	0	0.4725	0	0	0.1667	0.1591	1	0.4812	1
310011	0	0.142	0.4055	1	0.2097	0.5	0.2614	0	0.1073	0
310012	0	0.114	0.5153	0	1	0.1667	0.2727	0	0.2927	0
310013	0	0.05	1	0	0	0.3333	0.0568	0	0.3133	0
310014	0	0	0.8952	0	0.1814	0.4375	0.3409	0	1	0
310015	0	0	0.1741	0.1052	0.2299	0.25	0.0000	0	0.4618	0
310016	0	0	0.5968	0.5	0.236	0.25	0.4432	0	0.984	1
310017	0.5	0.402	0.2984	0.6052	0.2178	0.3125	0.0568	0	0.3956	0
310018	1	0	0.2984	0	0.1653	0.2917	0.0000	0	0.4759	1

表2　分类预测表

	观察值		预测值		
			Y		预测准确率
			0	1	
Step 1	Y	0	13	0	100
		1	3	2	40
	总预测准确率				83.33333
Step 2	Y	0	13	0	100
		1	2	3	60
	总预测准确率				88.88889
Step 3	Y	0	13	0	100
		1	2	3	60
	总预测准确率				88.88889
Step 4	Y	0	12	1	92.30769
		1	2	3	60
	总预测准确率				83.33333
Step 5	Y	0	13	0	100
		1	1	4	80
	总预测准确率				94.44444
Step 6	Y	0	12	1	92.30769
		1	2	3	60
	总预测准确率				83.33333

表3　进入方程的变量

		系数值（B）	标准误（S.E.）	卡方值（Wald）	自由度（df）	P值（Sig.）	OR值 Exp（B）
Step 1	论文被引	4.22	2.93236	2.0716	1	0.150069	68.06879
	常量	−1.6	0.73306	5.0088	1	0.025218	0.19386
Step 2	专利	3.2	2.2629	2.0059	1	0.156686	24.65455
	论文被引	5.87	3.80829	2.3782	1	0.12304	355.2851
	常量	−2.3	1.01384	5.2695	1	0.021702	0.097558
Step 3	专利	4.29	2.62142	2.6754	1	0.101909	72.80425
	论文被引	7.06	4.48256	2.481	1	0.115227	1165.155
	被专著综述引用	4.81	4.07276	1.3926	1	0.237965	122.2691
	常量	−3.2	1.35595	5.3971	1	0.020171	0.042848
Step 4	专利	6.14	3.69847	2.7564	1	0.096864	464.2279
	课题	4.51	4.01096	1.266	1	0.260521	91.19274
	论文被引	7.3	4.08105	3.199	1	0.073685	1479.148
	被专著综述引用	7.43	9.77798	0.5771	1	0.447466	1682.117
	常量	−6	3.46365	3.0408	1	0.081196	0.002382
Step 5	专利	6.4	4.36017	2.1539	1	0.142205	601.3064
	课题	5.18	4.87522	1.1305	1	0.287667	178.323
	项目执行时间	−6.6	9.80191	0.4521	1	0.501362	0.001374
	论文被引	6.83	4.05979	2.8295	1	0.092547	924.2863
	被专著综述引用	9.72	48.4507	0.0403	1	0.840984	16661.78
	常量	−4.1	3.91207	1.1138	1	0.29126	0.016105
Step 6	专利	5.25	3.50152	2.2452	1	0.134026	189.9411
	课题	4.67	4.04197	1.3322	1	0.248412	106.1971
	项目执行时间	−12	10.4569	1.3284	1	0.249086	5.83E−06
	论文被引	5.88	3.33023	3.1175	1	0.077454	357.8217
	常量	−1.4	2.52475	0.311	1	0.577065	0.244633

表4　预测分类表

		观察值	预测值		
			Y		预测准确率
			0	1	
Step 1	Y	0	16	0	100
		1	0	2	100
	总预测准确率				100

表5　进入方程的变量

		系数值（B）	标准误（S.E.）	卡方值（Wald）	自由度（df）	P值（Sig.）	OR值Exp（B）
Step 1	论文被引	317	46196.07611	4.7E−05	1	0.99453	4E+137
	常量	−124	17711.54995	4.9E−05	1	0.994417	1E−54

4.1　方法一后向逐步判别分析结果

（1）分类结果：后向逐步判别分析的预测准确率为88.88889%，有一项未授奖和一项授奖的项目评价预测有误（表6）。

（2）变量纳入结果：三步分析之后，排除了被检索工具收录、课题和参编专著；最终纳入的变量有：专利、会议交流、论文得分、论文被引、被专著综述引用和项目执行时间（表7）。

4.2　方法二后向逐步判别分析结果

（1）分类结果：后向逐步判别分析共进行了三步，最后的预测准确率为100%（表8）。

（2）变量纳入结果：三步判别分析之后，排除

表6　分类矩阵

预测准确率		G_1:0	G_2:1
		P=0.72222	p=0.27778
G_1:0	92.30769	12	1
G_2:1	80	1	4
Total	88.88889	13	5

注：列为预测分类；行为观察分类。

表7　纳入模型中的指标

	威尔克斯入	偏威尔克斯入	F-剔除	p-水平	容许度	拟合系数
专利	0.538450	0.772297	3.243219	0.099168	0.749023	0.250977
会议交流	0.435417	0.955047	0.517762	0.486805	0.906602	0.093398
项目执行时间	0.489832	0.848952	1.957146	0.189378	0.828441	0.171559
论文被引	0.594904	0.699010	4.736542	0.052195	0.775822	0.224178
被专著综述引用	0.548070	0.758742	3.497680	0.088283	0.808492	0.191508
论文得分	0.489285	0.849900	1.942693	0.190894	0.859403	0.140597

注：威尔克斯入：0.41584 approx. F(6,11)=2.5754 p<0 .0045。

表8　分类矩阵

预测准确率		G_1:0	G_2:1
		p=0.88889	p=0.11111
G_1:0	100	16	0
G_2:1	100	0	2
Total	100	16	2

注：列表示观察分类；行表示预测分类。

了专利、会议交流和参编专著；纳入模型的指标有：课题、被检索工具收录、项目执行时间、论文得分、被专著综述引用和论文被引（表9）。

以上两种方式用两种方法分析之后，结果与实际评奖结果相比较，将预测准确率和纳入变量的总结果以简表的形式表示（表10）。

5　讨论

5.1　关于Logistic回归与判别分析的结果

在预测结果方面：Logistic回归的预测准确率较判别分析的预测准确率略高。这可能是因为：判别分析要求样本最好是接近多元正态分布，且各组协方差矩阵相等。Logistic回归对样本的分布没有严格要求。

Logistic回归与判别分析要求变量间相互独立，无共线性，而样本中论文得分，论文被引和被检索工具收录之间存在一定的共线性关系，因而，在纳入变量方面，这两种方法均未纳入全部拟定的反映成果水平的自变量。变量间独立性问题在判别分析中的影响不是特别严重，这是因为判别分析关心的重点是对应变量的判别效果，而不是自变量的影响程度，因此，判别分析纳入变量较多。

5.2　关于方法一和方法二的评价结果

从预测准确率来看，方法二的预测准确率较方法一略高，说明在本研究的18个基础医学类项目中，一等奖项目在各项指标中的表现比非一等奖项目具有突出的优势。方法一和方法二这两种思维方式在实践中哪一种更具有优势还有待以后大样本研究进一步证实。建议在实际评价活动中，综合运用两种评价思维方式。在学术评价领域，永远没有完美指标，对任何指标的倚重总会使被评价者产生投机行为。

表9　纳入模型中的指标

	威尔克斯入	偏威尔克斯入	F-剔除	p-水平	1-容许度	拟合系数
课题	0.261492	0.865205	1.71375	0.217186	0.628995	0.371005
被检索工具收录	0.261969	0.863632	1.73690	0.214318	0.726337	0.273663
项目执行时间	0.239143	0.946063	0.62713	0.445151	0.841012	0.158988
论文被引	0.836381	0.270504	29.66484	0.000202	0.798244	0.201756
被专著综述引用	0.246730	0.916972	0.99600	0.339727	0.829613	0.170387
论文得分	0.291125	0.777138	3.15451	0.103350	0.560465	0.439535

注：威尔克斯入：0.22624 approx. F (6,11)=6.2700 p<0 .0045。

表10　两种评价方法的评价结果

			预测准确率	纳入变量
方法一	Logistic回归	前向逐步法	94.44444%	专利、论文被引、课题、项目执行时间、被专著综述引用
	判别分析	后向逐步法	88.88889%	会议交流、专利、论文被引、论文得分、被专著综述引用、项目执行时间
方法二	Logistic回归	前向逐步法	100%	论文被引
	判别分析	后向逐步法	100%	论文被引、论文得分、被检索工具收录、被专著综述引用、课题、项目执行时间

6　结论

对本研究采用的样本来说，从预测准确率方面，Logistic回归比判别分析稍高，但两者都筛选出了经验认为重要的变量，较全面反映科研成果水平的各个方面，初拟评价指标体系合理，可用于评价基础医学科研成果。但两种评价方法因其自身局限性影响，对变量容纳性还不够理想。实际评价中，应该进一步应用更加灵活、多维的评价方法，并与同行评价结合进行，以期为同行评议提供参考。

参考文献

[1] 邱均平,嵇丽.美国《科学引文索引》与科学评价研究[J].科研管理,2003,24(4):22-28.

[2] 杜波.应用计算机进行科研成果综合评价[J].油田地面工程,1994,13(1):73-74.

[3] 李红梅,张燕.科学技术活动的定量分析方法和工具[J].情报杂志,2001,20(B06):83-85.

[4] 叶深溪,许为民.文献计量学在科研评价中的应用进展[J].图书馆论坛,2003,23(4):12-14.

[5] Ingwersen P,Larsen B,Rousseau R,et al.论文-引文矩阵及其推导的定量评价指标[J].科学通报,2001,46(8):700-704.

[6] 秦志强,科研成果:论文和专利[J].现代情报,2003,23(8):47-49.

[7] 马斌荣.医学统计学[M].3版.北京:人民卫生出版社,2001.

[8] 蒋国华.科研评价与指标[M].北京:红旗出版社,2000.

[9] 洪楠,侯军,李志辉.STATISTICA FOR WINDOWS统计与图表分析教程[M].北京:清华大学出版社,北方交通大学出版社,2002.