

· 科学论坛 ·

基于词频分析的中国科研伦理研究评价

方玉东^{1*} 常宏建² 陈越¹ 雷鸣³

(1. 国家自然科学基金委员会, 北京 100085; 2. 山东财经大学, 济南 250014;
3. 北京邮电大学, 北京 100876)

[摘要] “科研伦理”正逐渐成为学界研究的热点, 为了评价国内科研伦理研究现状, 本文以 CNKI 中国知网数据库为来源, 以 2000—2013 年发表的核心期刊论文、优秀硕士和博士论文为评价资源, 以“科研伦理”、“科技伦理”、“科学伦理”为关键词进行检索, 应用词频分析法进行统计分析, 得出国内科研伦理的主要研究人员、研究热点、趋势、研究阶段和专业领域分布等状况, 并立足国内研究现状提出了改进建议, 以期为以后的研究提供借鉴。

[关键词] 科研伦理; 科技伦理; 科学伦理; 词频分析; 研究评价
DOI:10.16262/j.cnki.1000-8217.2015.05.010

科学研究是一种牵涉到科研人员、科技辅助人员、课题资助者、受试者、社会公众、政策制定者等诸多活动主体的社会活动^[1], 一旦产生社会活动, 伦理问题也就随之而来。在经历过核辐射、生态污染等诸多教训之后, 自 20 世纪 40 年代开始, 对科学研究的伦理学思考逐步展开。近年来, 随着我国科学技术研究的快速发展, “科研伦理”也逐渐得到各界的广泛关注, 并出现了“科技伦理”、“科学伦理”等相关概念^[2-4]。尽管上述概念存在差别, 但是学者们对它们的界定存在诸多交叉, 从广义的角度来看, 上述概念的研究内容都聚焦于科学与道德的关系问题^[5]。下文若无特别说明, 都以“科研伦理”代表这三方面的内容。

本文以“科研伦理”、“科学伦理”和“科技伦理”为关键词, 以核心期刊论文、优秀硕士论文和博士论文作为核心评价资源, 利用相关软件和词频分析工具, 对科研伦理研究热点、研究层次和发展动向等问题进行评价, 针对该领域的研究现状提出建议。

1 数据来源和样本筛选

以“科研伦理”、“科学伦理”和“科技伦理”为关键词搜索 CNKI 中国知网数据库、万方数据库和维普数据库中的期刊论文, 时间限定为 2000—2013

年, 对应 3 个数据库的论文数量分别为 697 篇、367 篇和 583 篇(表 1)。

表 1 各数据库以对应关键词查询到的论文数(篇)

数据库	关键词			
	科研伦理	科学伦理	科技伦理	合计
知网数据库	19	34	644	697
万方数据库	22	56	289	367
维普数据库	35	108	440	583

注: 时间范围 2000—2013。

由表 1 的统计结果可知, CNKI 中国知网数据库包含主题相关论文最全, 因此本文选择 CNKI 中国知网数据库为调研数据库。本次评估资源是 2000—2013 年中国知网数据库中科研伦理研究的核心期刊论文、优秀硕士论文与博士论文, 评估样本的筛选步骤如下:

(1) 初步检索。分别通过关键词“科研伦理”、“科学伦理”和“科技伦理”进行检索, 结果显示“科技伦理”相关研究论文 644 篇, 其中核心期刊论文 249 篇; “科学伦理”相关研究论文 34 篇, 其中核心期刊论文 12 篇; “科研伦理”相关研究论文 19 篇, 其中核心期刊论文 4 篇。以“或”逻辑输入三者显示相关研究论文 696 篇, 核心期刊论文 265 篇, 剔除 10 篇无效论文(会议通知、期刊目录和征文通知等), 得到

收稿日期: 2015-05-08; 修回日期: 2015-07-02

* 通信作者, Email: fangyd@nsfc.gov.cn

255篇有效核心期刊论文。

(2) 检索优秀硕士和博士论文。输入关键词“科技伦理”检索到优秀硕士论文 95 篇、博士论文 6 篇,以“科学伦理”为关键词检索到 2 篇优秀硕士论文,以“科研伦理”为关键词检索有 1 篇优秀硕士论文、1 篇博士论文,以“或”逻辑输入上述三个关键词有 97 篇优秀硕士论文、7 篇博士论文。

(3) 评估样本筛选。利用中国知网提供的数据分析,结合使用 excel 筛选上述 255 篇核心期刊论文。结果显示,发表 5 篇及以上与主题相关的研究论文的核心期刊有 11 种,共计 105 篇,占 255 篇核心期刊论文的 41.18%。具体统计结果如表 2 所示。

依照上述的论文选取方案,筛选出具有代表性的 7 篇博士论文、97 篇优秀硕士论文和 255 篇核心期刊论文,共计 359 篇论文作为本次评估的样本集,构建中国科研伦理研究数据库。基于中国知网数据库,运用 Excel 和词频统计方法进行统计分析。需要说明的是,CNKI 数据库中的期刊论文和博士、硕士学位论文的浏览分组有所不同,如学位论文的分组浏览中不含有基金选项,而期刊论文的分组浏览中未含导师选项,所以下文的部分统计结果中不一定同时含有期刊论文和学位论文,在具体的统计表的下面都注明了数据来源。

2 评价指标设计

本文所构建的评价指标体系如图 1 所示,该评价指标体系部分借鉴了田时中等^[6]的方法,如研究年份、研究人员、关注主题、专业领域、研究阶段和研究经费。不同的是,在研究阶段部分,通过作者发

表 2 255 篇核心期刊论文发表现状

序号	期刊名	载文量	百分比/%*
1	自然辩证法研究	27	10.59
2	哲学动态	14	5.49
3	道德与文明	12	4.70
4	科学技术与辩证法	12	4.70
5	伦理学研究	7	2.75
6	科技管理研究	6	2.35
7	科学学研究	6	2.35
8	科技进步与对策	6	2.35
9	自然辩证法通讯	5	1.96
10	理论月刊	5	1.96
11	求实	5	1.96

* 百分比是以各个期刊载文量与 255 篇核心期刊论文相比所得

量、合作度与合著率和研究层次三个二级指标来反映“科研伦理”在国内所处的研究阶段,其中研究层次可直接由 CNKI 数据库中查询统计得到。下面对作者发文量,合作度与合著率两个指标的含义做进一步解释。

(1) 作者发文量。核心作者是对某一学科领域研究的发展有较大贡献的科研人员,对核心作者的测评有助于学科研究和学术期刊的发展^[7]。通过对“科研伦理”领域发表论文数量的统计,可以找出本领域的发表论文最多的作者,然后基于普赖斯定律(即科学家总人数开平方,所得到的人数撰写了全部科学论文的 50%)可以得出科研伦理领域的核心作者,分析统计数据能反映出所研究领域核心作者群形成与否,从而反映出该研究领域是否得到持续稳定的资助^[8]。

(2) 合作度与合著率^[9]。合作度指的是每篇文章的平均创作人数,它可以反映出某领域内文章的协作完成水平,间接地说明该领域的研究程度。合著率是指共同编写的文章数与全部文章数的比值。通常某研究领域论文的合著率愈大,合作度愈高,该领域研究的就越透彻。所以,合著率是测度某一领域研究情况的重要参数,同时,合著率在一定程度上也能反映出学科的交叉性、渗透性以及某一学科相关论文的研究层次。

3 研究结果分析

3.1 研究主体

(1) 研究年份。由表 3 的论文统计结果可以看出,“科技伦理”为关键词的论文在 2000 年已有 10 篇,而“科学伦理”到 2001 年才开始有 1 篇文献论述,“科研伦理”到了 2003 年开始有 2 篇文献综述,

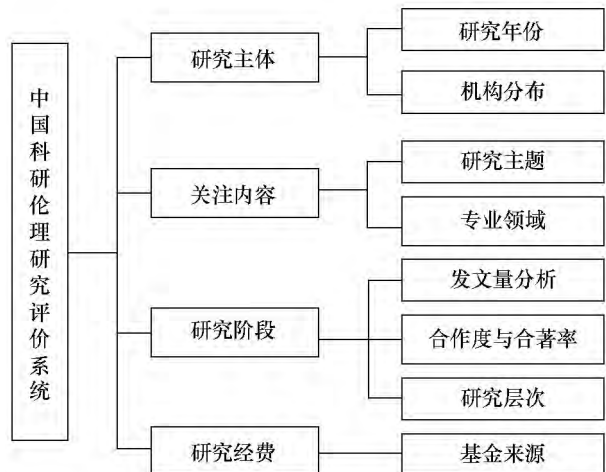


图 1 中国科研伦理研究评价指标体系

表 3 2000—2013 年中国科研伦理发表论文篇数统计

年度 关键词	00	01	02	03	04	05	06	07	08	09	10	11	12	13
科技伦理	10	12	21	20	23	24	25	28	37	29	22	30	42	27
科学伦理	0	1	0	2	1	1	0	2	1	1	0	1	2	2
科研伦理	0	0	0	2	0	0	0	0	1	0	0	0	2	1
合计	10	13	21	24	24	25	25	30	39	30	22	31	46	30

“科技伦理”的论文数量保持在 10—45 篇区间内，而“科研伦理”和“科学伦理”的文献数量则一直在 5 篇以下，“科技伦理”相对而言是该研究领域的热点话题。从总体变化趋势来看，科研伦理研究总的文献数量在 2000 年到 2008 年增加较为明显，总体呈现增加的趋势。

(2) 机构分布。应用频次统计方法分析数据发现，71.55% 的研究人员来自高等院校；其次是研究所/学会，占 1.6%；再次是科研管理机构，占 0.80%；另有 26.02% 并没有注明作者单位信息（见表 4）。可见，高等院校是科研伦理的主要研究群体。

3.2 关注内容

(1) 研究主题。为了解国内学者对“科研伦理”研究的关注重点，我们对 255 篇核心期刊论文中词频较高的关键词分组、合并（去除“科研伦理”、“科学伦理”和“科技伦理”关键词，再对相关的关键词进行合并），得到国内“科研伦理”研究所关注的重点，主要涉及到 14 个关键词（表 5）。从表 5 可以看出中国科研伦理研究的重点是“科学问题”、“社会问题”、“科技问题”和“哲学问题”，其比例分别占 23.50%、15.69%、10.98% 和 6.27%。科学问题包括科学研究、科学活动实践、科学精神与学术自由，社会问题包括社会责任、风险、和谐与创新，科技问题包括科技发展、科技活动、科技管理、人本科技、科技创新和科技责任等，哲学问题包括道德哲学、马克思主义哲

学、文化哲学、技术哲学、科技哲学和应用哲学等。科技人才与大学生是被关注的群体，其他方面有环境、经济、生命等学科。

(2) 专业领域。从 CNKI 知网数据库对于学科的分组浏览统计显示，国内“科研伦理”研究人员的专业分布相对广泛，主要包含的学科门类有：理学、经济学、哲学、教育学、文学、管理学、医学、伦理学和工学等。统计结果显示，伦理学所占比重最大，为 59.05%，而哲学、教育学、管理学、理学也占较大比重，专业领域分布如表 6 所示。

3.3 研究阶段

(1) 发文量分析。对以第一作者身份发表论文章的数量进行统计，东南大学科学技术伦理学研究所陈爱华发表的论文篇数最多，发表 7 篇。依照普赖斯定律核心作者计算公式 $m = 0.749(n_{max}^{1/2})$ ，(n_{max} 表示在指定期间内某一作者的最大发文量，在本文中 $n_{max} = 7$)，计算得出在本文中 $m = 1.98 \approx 2$ 。依照普莱斯定律，科研伦理研究领域的核心作者是发表论文章数不少于 2 篇的作者^[7]。表 7 统计了以第一作者身份发表不少于 2 篇论文的作者。

表 4 作者机构分布

作者单位	高等院校	研究所/学会	科研管理机构	未注明
频次	264	6	3	96
百分比/%	71.55	1.6	0.8	26.02

注：博、硕士论文的研究机构是高等院校

表 5 中国科研伦理研究关键词类分布

关键词类	科技人才	科学问题	科技问题	工程伦理	哲学问题	可持续发展	大学生
频次	2	60	28	6	16	2	4
百分比/%	0.78	23.50	10.98	2.35	6.27	0.78	1.59
关键词类	政治伦理	伦理困境	人文精神	辩证法	科学发展观	网络伦理	社会问题
频次	5	4	6	4	5	5	40
百分比/%	1.96	1.59	2.35	1.59	1.96	1.96	15.69

注：包含期刊论文和学位论文在内

表6 专业领域分布

专业领域	频次	百分比/%
伦理学	212	59.05
哲学	38	10.59
自然科学理论与方法	37	10.31
科学研究管理	33	9.19
高等教育	19	5.29
马克思主义	7	1.95
医学教育与医学边缘学科	5	1.39
工业通用技术及设备	4	1.11
经济理论及经济思想史	3	0.84
新闻与传媒	3	0.84
企业经济	3	0.84
职业教育	3	0.84
世界文学	3	0.84
教育理论与教育管理	3	0.84
心理学	2	0.56
社会学及统计学	2	0.56
生物学	2	0.56
其他(频次为1)	18	5.01

注:包含期刊论文和学位论文在内,其他专业的频次总和为18(每个专业的频次为1)

表7 第一作者身份发表不少于2篇的作者

序号	第一作者	发文篇数	序号	第一作者	发文篇数
1	陈爱华	7	14	李侠	2
2	孙道进	5	15	梁红秀	2
3	洪晓楠	3	16	刘魁	2
4	陈万求	3	17	彭利林	2
5	王雨辰	3	18	史兆光	2
6	张岂之	3	19	王一平	2
7	牛俊美	3	20	吴翠丽	2
8	段伟文	2	21	薛建明	2
9	冯昊青	2	22	张春美	2
10	李桂花	2	23	张澍军	2
11	李建会	2	24	杨怀中	2
12	李锐锋	2	25	张运松	2
13	李三虎	2	26	贺汉魂	2

注:以255篇核心期刊论文为数据库的统计结果

统计结果显示,论文篇数大于2(含2)的人数有26名,共计发表65篇,占255篇核心期刊论文的25.49%。根据普赖斯定律,核心作者的论文篇数是论文总量的一半时(即50%),才算该研究领域形成了核心作者群,而本文分析得到的25.49%仅为普赖斯定律的50%的一半略强,反映了形成科研伦理

的核心作者群还未形成。

(2) 合作度与合著率分析。在255篇核心期刊论文中,提供作者的论文有254篇,涉及作者总量为336人次。论文合作度是1.323(合作度为某种期刊在一定时期内作者总数与论文总数的比值),即每篇论文平均由1.323人共同完成。有73篇论文是由不少于两个作者共同完成的,合著率是28.74%,(其中67篇论文是由2人共同完成,合著率是26.38%;4篇论文由3人合著,合著率为1.57%)。尤其是有180篇论文只有一个作者,独著率高达70.87%。低合著率、高独著率在一定程度上反映了作者之间缺乏交流。

(3) 研究层次。从表8所示的研究层次分布的统计结果来看,国内科研伦理研究期刊论文和学位论文的研究层次基本一致,主要集中在“基础研究阶段”,包括基础研究(社科)和基础与应用基础研究(自科),期刊论文和学位论文分别占77.65%和84.62%。但基础研究(社科)所占的比例远高于基础与应用基础研究(自科),期刊论文和学位论文分别高出53.33%和78%。政策研究包括政策研究(社科)和政策研究(自科)在期刊论文和学位论文中所占的比重分别是17.26%和10.58%。其他包括行业指导、高等教育和工程技术等无论是期刊论文还是学位论文所占的比重都很小。

3.4 研究获基金资助情况

图2所示的基金资助情况的统计结果显示,255篇核心期刊论文中,222篇论文的研究没有标注基金资助,占87.06%。27篇论文获得国家级基金资助,仅占10.59%(其中绝大多数是由国家社科基金

表8 期刊论文和学位论文的研究层次分布

研究层次	期刊论文 /篇	百分比 /%	学位论文 /篇	百分比 /%
基础研究(社科)	167	65.49	83	79.81
政策研究(社科)	37	14.51	11	10.58
基础与应用基础研究(自科)	31	12.16	5	4.81
政策研究(自科)	7	2.75	0	0
行业指导(社科)	6	2.35	5	4.81
高等教育	5	1.96	0	0
文艺作品	1	0.39	0	0
工程技术(自科)	0	0	1	0.96

注:期刊论文的百分比是以255篇核心期刊论文为基数;学位论文的百分比是以97篇优秀硕士论文和7篇博士论文为基数。

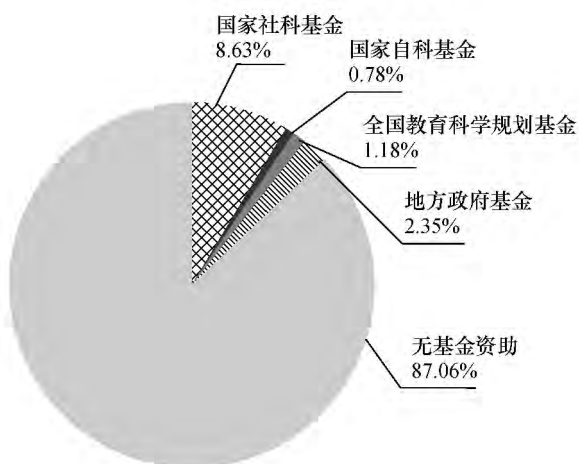


图2 研究获基金资助情况

资助的,国家自科基金资助的比重仅占 0.78%),地方政府基金资助的只有 6 篇,占 2.35%。

4 结论与建议

4.1 研究结论

通过对 2000—2013 年 CNKI 中国知网数据库中科研伦理研究的核心期刊论文、优秀硕士论文和博士论文进行统计分析,可以发现国内对于科研伦理的研究主要有以下 5 个特点:

(1) 研究主体集中在高等院校。本文研究结果表明,对国内科研伦理进行研究的人员主要来自高等院校。除去未注明作者单位信息的论文,71.55% 的研究人员来自高等院校,占了相当大的比重;其次是研究所/学会,占 1.6%;再次是科研管理机构,占 0.8%。研究表明,我国科研伦理研究的主体非常集中。

(2) 研究热点相对集中。国内科研伦理研究的热点研究反映了此学科的大方向。本研究结果显示,国内科研伦理研究的重点主要集中在“科学问题”、“社会问题”、“科技问题”和“哲学问题”四大主题上,尤其是科学问题,其在国内科研伦理研究的比例接近四分之一。这说明国内科研伦理核心主题比较集中。

(3) 国内科研伦理研究仍处在基础研究阶段。基础理论研究和实践效果评估是科学研究的不同层级,但都是服务于实现科学研究的目的是。对国内科研伦理研究的阶段分析表明,77.65% 的期刊论文和 84.62% 的学位论文处在“基础研究”阶段,即关注的还是基础理论的研究和探讨,缺乏对实践效果的评价,整体上还处在基础研究阶段。

(4) 核心作者群尚未形成,跨行业交流合作少。统计结果显示科研伦理研究领域核心作者发文量所占比重为 25.49%,远小于普赖斯定律的 50%,说明科研伦理的核心作者群还没有形成。从合作度和合著率来看,独著率高达 70.87%,在一定程度上反映出科研伦理领域的研究人员单打独斗现象明显、跨行业交流不多,不利于该领域的可持续深入研究。

(5) 缺乏研究资助。统计结果显示,有 87.06% 的研究没有基金资助,而获得国家级基金的只有 10.59%,地方政府基金只占 2.35%。资金是科研人员进行研究必不可少的保障,从课题前期调研至实验阶段,都离不开资金的支持。特别是一些调研类课题,没有一定的基金资助很难开展。过少的研究资助一方面降低了研究人员勇攀科研高峰的积极性,另一方面说明国家对于科研伦理问题的重视程度还有待加强。

4.2 相关建议

影响科研伦理研究质量的因素是多方面的,不仅受研究者的水平和研究机构的制度影响,也受政府管理部门的影响。协调好研究者、研究机构和政府部门之间的关系才能有效提高研究质量。

(1) 研究者。研究者的学术水平、学术规范直接影响了研究的质量。需要提高研究者的学术造诣,培养务实的精神,从主观上端正对待科学研究的态度。研究方法的使用直接影响到研究成果的优劣,这也是反映研究者学术水平、学术规范的重要指标,科研伦理研究运用的研究方法相对比较单一,势必对研究质量造成一定的影响,所以需要研究者掌握更多的方法,从而形成系统、全面的研究成果。

(2) 研究机构。研究者绝大多数来自高等院校,其次是研究所。研究机构的体制在很大程度上对研究者的研究方向具有导向作用。中国科研伦理的研究主题相对比较集中,研究机构可以充分利用这一有利的条件,对核心主题展开有针对性的研究,促使研究质量更上一层楼。另外研究机构应该提供有效合理的科学评价体系,为科研者的科学研究提供正反馈,使促进机制与主观积极性进入良性循环,以期达到更好的效果。

(3) 政府及企事业单位。缺乏基金资助很大程度上影响了科研伦理的研究进展。研究结果表明,有高达 87.06% 的研究没有基金支持,这也是导致国内科研伦理研究方法主要集中在文献分析、理论

分析而实证分析相对不足,缺乏跨行业交流合作的一个重要原因。政府是基金的直接控制者,各企事业单位也是研究基金的重要来源。应逐步形成以政府为主导,各企事业单位加大研究基金投入的格局,加大对科研伦理研究领域的投入,为研究机构的研究者提供良好的科研环境,从而促进研究质量的提高。

本文是以中国知网数据库为例对科研伦理进行的初步评价,尚存在一定的不足。譬如,本文所查询的文献仅以核心期刊论文、优秀硕士论文和博士论文作为核心评价资源,而对著作、会议论文和翻译作品等没有收集。实际上,除了论文类成果,社会科学的成果形式还包括专著和翻译作品等,其中专著应该纳入评价的范围中。在后续研究中,我们将对更为广泛的数据库进行查询,完善评价指标,以期作出更为客观的评价。

参 考 文 献

- [1] 陈文化,李立生.“科技伦理”是一种抽象的伦理观. 自然辩证法研究, 2011, 17 (9): 23—27.
- [2] 宋启林. 论科学伦理的多重维度. 自然辩证法研究, 2003, 19 (10): 76—80.
- [3] 刘国云. 中外科学伦理发展比较研究. 自然辩证法研究, 2013, 29 (2): 59—64.
- [4] 江雪莲. 论科技伦理学的研究特色. 自然辩证法研究, 2003, 19(9): 64—70.
- [5] 方玉东, 张莉莉, 陈越. 国外科研伦理管理理论与实践综述. 山东农业大学学报(社会科学版), 2011 (6): 97—101.
- [6] 田时中, 田家华, 张光进. 基于论文词频分析的中国低碳经济研究评价. 国土资源科技管理, 2012, 29 (1): 74—82.
- [7] 刘小兵. 基于CNKI数据库的国内民国文献研究的文献计量分析. 新世纪图书馆, 2013, 8: 29—34.
- [8] 钟文娟. 基于普赖斯定律与综合指数法的核心作者测评—以《图书馆》建设为例. 科技管理研究, 2012 (2): 57—60.
- [9] 邱均平, 温芳芳. 作者合作程度与科研产出的相关性分析. 科技进步与对策, 2011, 28 (5): 1—5.

Evaluation of scientific research ethics studies in China based on word frequency analysis

Fang Yudong¹ Chang Hongjian² Chen Yue¹ Lei Ming³

(1. National Natural Science Foundation of China, Beijing 100085; 2. Shandong University of Finance and Economics, Ji'nan 250014; 3. Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract As “scientific research ethics” is gradually becoming a research hotspot, “scientific research ethics”, “science and technology ethics”, and “scientific ethics” are set as keywords in this paper to search core journal papers, outstanding master’s and doctoral dissertations published from 2000 to 2013 in the China National Knowledge Internet (CNKI) database, in order to evaluate the current domestic situation of the studies on scientific research ethics. Word frequency analysis is used to statistically analyze the situations of the main researchers, research focuses, trends and stages and the field distributions of domestic studies on research ethics. Suggestions for improving the domestic research status is put forward so as to provide references for following-up studies.

Key words scientific research ethics; science and technology ethics; scientific ethics; word frequency analysis; research evaluation