

我国 CRO 研究热点分析*

任芮 褚淑贞

(中国药科大学国际医药商学院 南京 211198)

[摘要] 采用共词分析法分析我国 CRO 研究热点, 首先对高频关键词进行分类汇总, 归纳出 CRO 研究热点的微观和宏观两个视角; 其次构建共词矩阵进行聚类分析, 分别获得宏观和微观视角下的研究热点; 最后运用多维尺度分析, 得出产业链中研究热点集中于新药开发和临床试验环节。

[关键词] 合同研究组织; 医药产业; 生物医药; 临床试验; 新药开发

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2016.10.011

Analysis on Research Hotspots of CRO in China REN Rui, CHU Shu-zhen, International School of Medicine, China Medical University, Nanjing 211198, China

[Abstract] The paper analyzes the research hotspots of the Contract Research Organization (CRO) in China by co-word analysis. Firstly, it classifies high-frequency keywords and summarizes the microscopic and macroscopic perspectives for CRO research hotspots. Then, it builds the co-word matrix, makes cluster analysis and obtains respective research hotspots from macroscopic and microscopic perspectives. In the end, by multi-dimensional scaling analysis, it concludes that research hotspots in the industrial chain focus on the development of new drugs and clinical trials.

[Keywords] Contract Research Organization (CRO); Pharmaceutical industry; Biological medicine; Clinical trial; New drug development

1 引言

1.1 医药研发合同组织概述

我国本土医药企业长期以仿制药生产为主, 创新性较低。近年来, 随着国家不断强调研发创新的重要性, 我国医药研发企业也不断崛起, 但是相较于跨国企业而言, 还有很大的差距。医药研发合同研

究组织 (Contract Research Organization, CRO) 的出现有效降低医药研发成本, 加快新产品上市的速度, 缩短周期, 为产品赢得更多的上市时间, 并且带来更多的收益。因此, 越来越多的医药研发企业选择将研发全部或者部分研发环节外包给 CRO, 从而获得更高的研发效率^[1]。根据《药物临床试验质量管理规范》的定义, CRO 是一种学术性或商业性的科学机构, 申办者可委托其执行临床试验中的某些工作和任务, 此种委托必须做出书面规定。根据实际情况可以将 CRO 理解为在医药研发领域内的以合同形式提供部分或全部技术支持的第 3 方外包服务企业。

CRO 兴起于美国, 起初是数据处理和分析的咨

[修回日期] 2016-05-27

[作者简介] 任芮, 硕士研究生。

[基金项目] 江苏省普通高校学术学位研究生科研创新计划 (项目编号: 1150031412)。

询公司,最早可追溯到20世纪70年代,美国的一些研究机构为制药企业提供临床服务。20世纪90年代美国食品药品监督管理局正式批准CRO企业上市,逐渐形成了医药研发外包服务产业。2005年以后,CRO在我国部分地区逐渐建立和发展起来,至今已有10多年的成长历程。我国CRO主要集中于北京、上海、南京、无锡等城市并形成了一定规模的产业集群,如上海张江的国家上海生物医药科技产业基地——张江生物医药基地(张江药谷)、北京中关村生物医药园等。目前,我国本土CRO企业已有上千家,但是多以技术含量相对较低的注册服务为主,并且CRO企业的服务质量良莠不齐,很多发展较为成熟的CRO企业也被外资企业并购。虽然有像药明康德和泰格这样的领头企业,但是我国CRO的发展总体还处于初级阶段。因此,对于CRO企业的研究,一方面对于我国的医药研发至关重要,另一方面也关系到我国新型产业结构调整。

目前,由于数据的缺乏,实证研究相对较为困难,国内学者对于CRO的研究主要集中于理论性研究;也有学者运用博弈论的方法进行实践模拟,或者通过某一个企业的数据进行实证分析。笔者认为,研究思考医药研发合同组织的中国发展问题上,首先需要系统地汇总并分析目前学术界对医药研发合同组织的研究现状以及分析结果,通过对文献的统计分析,奠定研究医药研发合同组织的理论基础。

1.2 共词分析法

共词分析法被广泛应用于科学计量学、人文学科计算研究等领域。共词分析法是一种内容分析方法,通过统计一组词中每两个词语在同一文献中出现的频数,据此进行聚类分析,依据分析结果判断所研究问题的热点以及发展趋势^[2]。共词分析法通常被学者用来分析领域内的研究热点,一方面反映了学科的研究水平和发展动态;另一方面,揭示了学科之间的相互关系。相较于其他文本分析方法,共词分析方法更加灵活,结果也比较直观,便于分析。本文利用共词分析法,对CRO企业在我国的研究热点进行分析,科学地展现CRO企业的研究现状并探究发展趋势。

2 数据来源及处理

2.1 数据来源

以《中国学术期刊网络出版总库》为文献检索来源,对2006-2015共10年间的文献进行检索,选择其中的期刊文献作为研究对象,通过关键词的筛选,对文献进行计量学统计。

2.2 筛选文献和关键词

选择筛选出主题为药物研发合同研究组织或药物研发外包或篇名为CRO的文献,共计580条相关记录。为确保文献的有效性,剔除与主题无关或新闻等文献,筛选后共计有88篇文献^[3];将88篇文献进行格式转换,导入SATI软件,统计出关键词共325个。

2.3 关键词规范化

对325个关键词进行规范化处理。首先,删除与研究主旨不符的含有人名等的关键词,得到关键词292个。同时,考虑到CRO研究文献数量较少、文献研究比较分散、集中度低,剔除了出现频率仅1次的关键词200多个,经过关键词的分类汇总之后,获得有效关键词,即至少频次为2的关键词有54个,见表1。将54个关键词带入Donohue提出的区分高低频次词分界的公式:

$$T = (-1 + \sqrt{1 + 8 * I}) / 2$$

式中: $I=54$,所以 $T=10$ 。高于10的关键词有CRO、研发外包、药明康德、临床试验、医药研发和生物医药6个。由于“CRO”、“研发外包”和“药”是在文献检索时既已设定的主题,因此将CRO、研发外包和医药研发归为一类汇总。初步分类汇总后,符合条件的高频词有CRO、药明康德、临床试验和生物医药4个,明显不能全面表现出我国学者对CRO的研究。另外,从表1中可以发现,很多关键词都存在相近或同义的关系,如生物医药和生物制药^[4]、制药工业和医药产业等,因此,在去除研发等相关词类后,汇总出18个高频词,见表2。

表 1 关键词词频情况

关键词	频次	关键词	频次	关键词	频次
CRO	48	制药工业	3	企业全面风险管理	2
研发外包	24	医药产业	3	新药	2
药明康德	13	制药企业	3	专利保护期	2
临床试验	12	制药产业	3	阿斯利康	2
医药研发	12	全球化学	2	药品开发	2
生物医药	11	生物医药科技	2	药物临床试验	2
合同研究组织	9	临床研究	2	科技产业基地	2
医药企业	8	药物临床研究	2	中国医药产业	2
制药公司	8	实验室研发	2	化学药	2
创新药物	6	公司合并	2	创新药	2
药品研发	6	医药产业基地	2	研发	2
生物技术公司	5	研发平台	2	研发领域	2
张江	5	对策	2	策略	2
新药研发	5	新兴产业	2	生物医药企业	2
新药研究开发	4	人才流失	2	研发服务	2
大型制药企业	4	生物医药公司	2	新药创制	2
生物制药	3	医药行业	2	跨国制药企业	2
新药研究	3	风险评估	2	仿制药	2

表 2 高频关键词分类汇总

关键词	频次	关键词	频次	关键词	频次
生物医药 (生物医药、生物医药科技等)	25	产业基地 (张江、科技产业基地等)	10	ERP 实施	2
临床试验 (临床试验、药物临床试验等)	18	策略分析 (策略、对策等)	4	专利保护	2

表 3 关键词相似矩阵 (部分)

关键词	CRO	研发外包	药明康德	临床试验	医药研发	生物医药	合同研究组织	医药企业	制药公司	药品研发
CRO	1	0.109 6	0.081 9	0.181 2	0.088 8	0.071 1	0.002 4	0.067 9	0.067 9	0.058 0
研发外包	0.109 6	1	0.320 5	0.013 9	0.281 2	0.136 4	0.004 6	0.046 9	0.130 2	0.062 5
药明康德	0.081 9	0.320 5	1	0.025 6	0.102 6	0.062 9	0	0.086 5	0.086 5	0.115 4
临床试验	0.181 2	0.013 9	0.025 6	1	0.006 9	0.068 2	0.009 3	0.093 8	0.093 8	0.055 6
医药研发	0.088 8	0.281 2	0.102 6	0.006 9	1	0.068 2	0	0.166 7	0.093 8	0.125 0
生物医药	0.071 1	0.136 4	0.062 9	0.068 2	0.068 2	1	0	0.181 8	0	0
合同研究组织	0.002 4	0.004 6	0	0.009 3	0	0.181 8	1	0	0	0
医药企业	0.067 9	0.046 9	0.086 5	0.093 8	0.166 7	0	0	1	0	0.187 5
制药公司	0.067 9	0.130 2	0.086 5	0.093 8	0.093 8	0	0	0	1	0.020 8
药品研发	0.058 0	0.062 5	0.115 4	0.055 6	0.125 0	0.060 6	0	0.187 5	0.020 8	1

续表 2

药企 (医药企业、大型医药企业等)	17	风险评估 (企业全面风险管理等)	4	全球化学	2
药明康德	13	发展前景	2	阿斯利康	2
医药产业 (制药产业、中国医药产业等)	12	新兴产业	2	桑迪亚	2
制药公司 (公司合并等)	10	人才流失	2	实验室研发	2

根据表 2 的关键词词频分布初步分析,我国医药 CRO 的相关研究涉及了医药行业、产业研究,如生物医药、临床试验等;也涉及了医药企业、公司的相关研究,如风险评估、策略分析等。可以初步观察出我国 CRO 的相关研究既包含了宏观研究也包含了微观分析。为了更加全面、具体地分析 CRO 相关研究热点,运用 SPSS 软件对数据进行统计分析。

3 统计分析

3.1 矩阵构建和转化

在进行计量分析前,需要将共词矩阵转化为相关矩阵,引入 Ochiai 系数^[5]:

$$\text{Ochiai 系数} = \frac{\text{两个关键词的共词频数}}{\sqrt{\text{两个关键词的频次乘积}}}$$

计算出每个对应 Ochiai 系数后,得到高频关键词 54 × 54 的相关矩阵,见表 3。

Ochiai 系数在 $[0,1]$ 区间内, 数值越接近 1, 表示两个关键词之间关联性越大; 反之, 数值越小, 则表示关键词之间的差异性越大。

3.2 聚类分析

由于 54 个变量数量较大, 并且词义上存在部

分重复性, 因此, 对分类汇总后的关键词进行聚类分析。采用 EXCEL 汇总表 2 中 18 个高频关键词的共词矩阵, 再去企业名称的数列, 形成 15×15 的矩阵, 见表 4。

表 4 高频关键词相似矩阵

关键词	临床	生物医药	药企	制药公司	医药产业	策略分析	产业基地	发展前景	新兴产业	人才流失	风险评估	ERP 实施	专利保护期	全球化学	实验室研发
临床	1	0.235 7	0.285 8	0.298 1	0	0	0.074 5	0.333 3	0	0	0.235 7	0	0.166 7	0	0
生物医药	0.235 7	1	0.436 6	0.316 2	0.173 2	0.3	0.3162	0.141 4	0.565 7	0	0.2	0	0.141 4	0	0.282 8
药企	0.285 8	0.436 6	1	0.153 4	0.21	0	0.460 2	0.171 4	0	0	0	0	0.171 5	0	0.171 5
制药公司	0.298 1	0.316 2	0.153 4	1	0.091 3	0	0	0	0.223 6	0	0	0	0	0.223 6	0.223 6
医药产业	0	0.173 2	0.21	0.091 3	1	0	0.182 6	0	0.204 1	0	0	0	0.204 1	0.204 1	0
策略分析	0	0.3	0	0	0	1	0	0	0	0	0	0	0	0	0
产业基地	0.074 5	0.316 2	0.460 2	0	0.182 6	0	1	0	0	0	0	0	0	0	0
发展前景	0.333 3	0.141 4	0.171 4	0	0	0	0	1	0	0	0	0	0	0	0
新兴产业	0	0.565 7	0	0.223 6	0.204 1	0	0	0	1	0	0	0	0	0	0
人才流失	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
风险评估	0.235 7	0.2	0	0	0	0	0	0	0	0	1	0	0	0	0
ERP 实施	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
专利保护期	0.166 7	0.141 4	0.171 5	0	0.204 1	0	0	0	0	0	0	0	1	0	0
全球化学	0	0	0	0.223 6	0.204 1	0	0	0	0	0	0	0	0	1	0
实验室研发	0	0.282 8	0.171 5	0.223 6	0	0	0	0	0	0	0	0	0	0	1

选择 Z 得分, 得到聚类树状图。进一步将关键词分成 5 类: 第 1 类, 药企、产业基地和生物医药; 第 2 类, 临床、风险评估和发展前景; 第 3 类, 策略分析、人才流失、ERP 实施; 第 4 类, 医药产业、专利保护期; 第 5 类, 制药公司、全球化学、新兴产业和实验室研发^[6], 见图 1。在高频关键词的观察基础上, 通过系统聚类分析, 具体分析我国

CRO 研究热点。从宏观角度分析, 我国 CRO 研究热点集中于探索生物医药产业园区的发展现状、从全球视角来分析全球实验室的建立、药品专利对我国 CRO 的影响; 从微观角度分析, 我国 CRO 的研究热点集中于对临床型企业的风险管理和发展分析、企业的人力资源和发展策略管理^[7]。

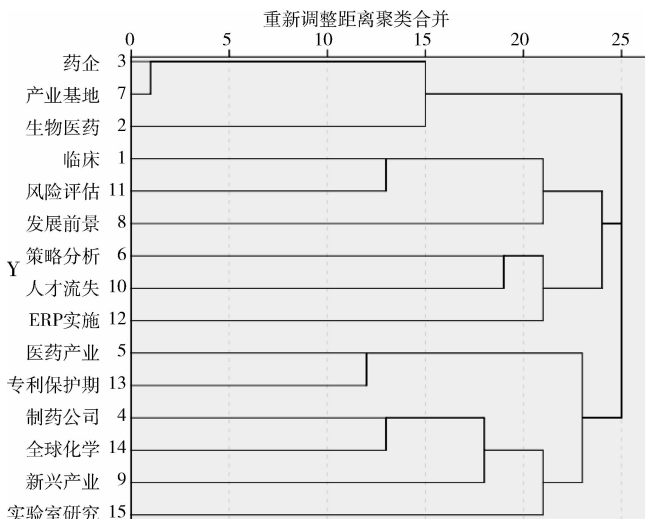


图 1 聚类分析

3.3 多维尺度分析

为了避免在归类汇总中后聚类分析产生的误差，对 54 个关键词进行多维尺度分析，更加直观地展现 CRO 研究关键词的分布。直接将表 3 数据导入，选择多维尺度分析的 Alscal 方法，创立二维感知图；选择“从数据创建距离”、度量模型为 Euclidean 距离，最后得到图 2。

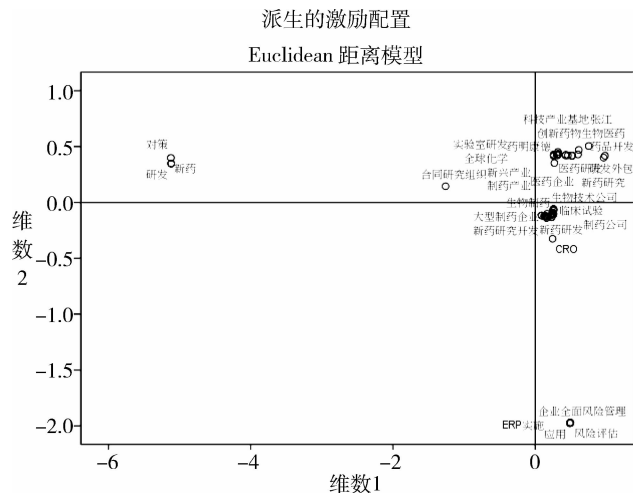


图 2 多维尺度分析

Stress 系数为 0.180 22，RSQ 值为 0.953 32，压力系数在 0.1~0.2 之间属于一般水平，RSQ 值在 0.6 以上且越高数据越理想，0.953 32 充分满足统计学要求。由图 2 可知：横向维度观察，左侧主要是全球视野下的产业研究，右侧是生物医药产业

园区的临床试验企业研究，因此将横向维度定义为产业-企业维度；从纵向维度上分析，上端主要分布着医药产业和企业，下端分布着企业管理中的风险评估、ERP 实施等^[8]，因此定义纵向维度为医药研发-管理维度。横向维度的定义与高频关键词观察所得结果一致，我国 CRO 热点研究既有宏观行业的研究也有微观企业的相关探究；多维尺度分析中纵向维度的定义提供了不同于其他方法研究得出的结果，CRO 的相关研究在涉及医药研发时重点在于新药的开发和临床试验环节，在讨论企业问题中主要探究企业风险管理和资源利用问题。从空间分布特点来看，大部分关键词都在中心集聚，通常说明国内学者对 CRO 的相关研究比较集中；但是，由于这些关键词是进行了初步筛选后的，已经剔除了频数为 1 的关键词，因此，只能从集中度来看，依旧可以观察到国内 CRO 研究热点有宏观的产业研究，也有微观的企业研究，其中对于生物医药领域的 CRO 研究较为突出，总体与前文词频汇总观察和聚类分析的结果相一致。

4 结语

本文利用高频关键词的统计汇总，初步总结出我国 CRO 相关研究热点分为两类：一类是医药产业领域的宏观角度研究，另一类是医药企业领域的微观角度研究。运用 SPSS 对共词矩阵进行聚类分析和多维尺度分析，通过聚类分析归纳出我国 CRO 研究宏观角度下的热点集中于探索生物医药产业园区的发展现状、从全球视角来分析全球实验室的建立、药品专利对我国 CRO 的影响等；在微观角度下的热点集中于对临床型企业的风险管理和发展分析、企业的人力资源和策略管理等。最后，通过多维尺度分析，对 CRO 研究热点进行了维度分析和分布点分析，维度分析中提出了研发产业链中新药发现和临床试验环节是文献研究的热点领域。

参考文献

1 鲁菁. 中国医药研发外包服务产业发展研究 [D]. 长沙: 中南大学, 2012: 34-62.

(下转第 61 页)

入指标很难界定,目前的实证研究只能采取可获指标进行分析。本文研究通过影响医院科研活动的因素确定科研效率的指标体系,运用主成分分析法对研究数据进行初步处理,并用随机前沿分析方法进行医院科研活动效率评价,通过 20 家三甲医院数据进行实证分析,应用结果说明了模型的可行性,并得出目前科研项目投入仍然是我国医院科研效率提高的一个关键因素,为进一步合理规划医院科研资源提供依据。

参考文献

- Aigner D J, Lovell C A K, Schmidt P. Formulation and Estimation of Stochastic Frontier Production Function Models [J]. *Journal of Econometrics*, 1977, 6 (1): 21 - 37.
- Battese G E, Corra G S. Estimation of a Production Frontier Model with Application to the Pastoral Zone of Eastern Australia [J]. *Australian Journal of Agricultural Economics*, 1977, 21 (3): 169 - 179.
- Rosko M D, Mutter R L. Stochastic Frontier Analysis of Hospital Inefficiency - A Review of Empirical Issues and an Assessment of Robustness [J]. *Medical Care Research and Review*, 2008, 65 (2): 131 - 166.
- Linna M. Measuring Hospital Cost Efficiency with Panel Data Models [J]. *Econometrics and Economics*, 1998, 7 (5): 415 - 427.
- Gerdtham U G. Internal Markets and Health Care Efficiency: A Multiple - Outputstochastic Frontier Analysis [J]. *Health Economics*, 1999, 8 (2): 151 - 164.
- 吴明, 李曼春, 侯建林, 等. 随机前沿成本函数方法在医院经济效率评价中的应用 [J]. *中华医院管理杂志*, 2000, 16 (8): 507 - 509.
- 姚红, 胡善联, 曹建文. 上海市 45 家医院供给的技术效率评价 [J]. *中国医院管理*, 2003, 5 (5): 9 - 11.
- 王伟成, 曾武. 随机前沿成本模型在中医院技术效率评价中的应用 [J]. *中华医院管理杂志*, 2005, 21 (5): 333 - 336.
- 刘启贵, 宋桂荣, 禹海波, 等. 随机前沿方法在评价医院效率中的应用 [J]. *中国卫生统计*, 2005, 22 (5): 303 - 305.
- 李清华, 郭耀煌. 主成分分析用于多指标评价的方法研究——主成分评价 [J]. *管理工程学报*, 2002, 16 (1): 39 - 43, 3.
- Battese G E, Coell T J. A Model of Technical Inefficiency Effects in Stochastic Frontier Production for Panel Data [J]. *Empirical Economics*, 1995, (20): 40 - 43.
- 王建宏, 陆志峰, 高岩波, 等. 基于 PCA - DEA 和 PCA - SFA 的大型医院绩效评价 [J]. *数学的实践与认识*, 2011, 41 (15): 36 - 44.
- 国家卫生和计划生育委员会. 医院名录查询 [EB/OL]. [2016 - 03 - 06]. <http://www.nhfp.gov.cn/>.
- 邱均平. 信息计量学 [M]. 武汉: 武汉大学出版社, 2007.
- 袁晓园, 华薇娜. 基于共词分析的我国医学信息学国际发文热点研究 [J]. *医学信息学杂志*, 2014, 35 (4): 8 - 14.
- 赵健. 生物医药研发外包初探 [J]. *经营管理者*, 2014, (1): 113 - 114.
- 张翔, 王璨. 国内科技企业公司治理研究的主题分布及热点——基于共词分析法 [J]. *沈阳大学学报*, 2014, 16 (1): 14 - 19.
- 薛薇. 基于 SPSS 的数据分析 [M]. 北京: 中国人民大学出版社, 2006.
- 宋茂海, 李东方. 基于共词分析的国内生物信息学热点领域研究 [J]. *生物信息学*, 2014, 12 (1): 46 - 51.
- 刘跃. 我国 CRO 公司与国际 CRO 公司的管理对比分析 [J]. *中国卫生产业*, 2015, (23): 193 - 195.

(上接第 56 页)