

基于内容分析法的乳腺癌研究热点分析

张惠荣 姜玉梅

宋丽萍

(天津医科大学附属肿瘤医院图书馆 天津 300060) (天津师范大学管理学院 天津 300387)

〔摘要〕 应用内容分析法对乳腺癌研究热点进行分析,通过 Web of Science 平台的文献分析功能,确定 2005 - 2009 年乳腺癌领域高影响力的国家、机构和作者,对其所发表的被引频次最高的论文及领域内被引频次最高的论文进行内容分析与主题统计,确定近年来乳腺癌研究热点。

〔关键词〕 内容分析法;引文分析;乳腺癌

Analysis on Research Hotspots of Breast Cancer Based on Content Analysis Method ZHANG Hui - rong, JIANG Yu - mei, Library of Cancer Hospital Affiliated to Tianjin Medical University, Tianjin 300387, China; SONG Li - ping, College of Management, Tianjin Normal University, Tianjin 300387, China

〔Abstract〕 Analyzing breast cancer research hotspots applying content analysis method, through literature analysis function of Web of Science, the paper clarifies the countries, institutes and authors with high influence in breast cancer area from 2005 to 2009, carries out content analysis and subject statistics on the most highly cited papers published by the above countries, institutes and authors and the most highly cited papers in this area, so as to determine the research hotspots in breast cancer area.

〔Keywords〕 Content analysis method; Citation analysis; Breast cancer

内容分析法发源于传播学研究,在政治、经济、新闻、军事等社会学学科领域应用广泛^[1]。医学相关学科热点研究可以为临床工作者的科学决策提供参考,以便更好地进行临床实践,还可以为相关科研人员确定研究方向和研究内容提供依据。利用内容分析法可以客观地从宏观的角度来研究学科热点,分析学科发展过程与现状,并预测发展趋势^[2]。本文尝试将内容分析法应用于医学领域,以分析乳腺癌研究热点为例。

以 SCIE (SCI - Expanded) 数据库为资料来源,应用 Web of Science 平台的文献分析功能,结合 h 指数确定 2005 - 2009 年乳腺癌领域高影响力的国家、机构和作者,利用 2005 - 2009 年高影响力国

家、机构、作者所发被引频次最高的论文以及领域内被引频次最高的 15 篇论文,逐篇进行内容分析并进行主题类目统计,确定近年来乳腺癌热点。按论文发表时间顺序理清热点主题发展脉络,并通过高被引论文及阅读国内外综述性文章阅读研究各热点主题发展现状和趋势。

1 样本选择

本研究采用《中国科技论文统计报告》样本选取法进行样本选择,4 类论文包括: Article, Review, Letter 以及 Editorial Material。以乳腺癌相关研究论文为分析对象,检索方法为高级检索模式下输入检索式: TI = ((breast SAME cancer) OR (breast SAME carcinoma) OR (breast SAME tumor) OR (breast SAME tumour) OR (breast SAME neo-

〔修回日期〕 2010 - 06 - 18

〔作者简介〕 张惠荣,硕士,馆员,发表论文 8 篇。

plasms)) OR TI = ((mammary SAME cancer) OR (mammary SAME carcinoma) OR (mammary SAME tumor) OR (mammary SAME tumour) OR (mammary SAME neoplasms))，出版年选择 2005 – 2009，数据库更新时间截止到 2010 年 1 月 20 日。

2 建立类目体系

使用《中图法》^[3]肿瘤类的分类标准建立乳腺癌热点研究的类目体系，有利于确定规范的分析单元的归类标准。首先，初步阅读乳腺癌相关论文，对基本内容进行整体把握后，确定分析单元，将分析单元纳入到类目体系中。包含分析单元的类目体系，见表 1。

表 1 乳腺癌热点研究类目体系		
一级类目	二级类目	下位主题
流行病学与预防	易感基因	-
病理学病因学	病理生理学—病因学、发生学	内分泌因素—雌激素受体研究 生物因素—microRNA
	病理类型	分子病理学 基因表达谱预后 三阴乳腺癌
诊断学	超声诊断	-
	实验室诊断	肿瘤标志物
	放射线、同位素诊断	分子影像
治疗学	内分泌治疗	-
	综合疗法	-
	保守治疗	-
	化疗	-
	免疫治疗	-
	基因靶向治疗	-
	放射疗法	-
	手术（外科）疗法	-

3 2005 – 2009 年乳腺癌研究热点分析

3.1 高影响力国家及其被引频次最多的论文主题分布

利用 Web of Science 的分析功能，对发文国家进行分析，按乳腺癌相关研究领域发文量进行排序，并对各国发表的论文从篇数、篇均被引频次、

h 指数、单篇被引最高频次几方面进行分析，见表 2。

表 2 发表乳腺癌相关论文数量居前 10 位的国家			
国别	论文量 (篇)	篇均被 引频次	h 指数
美国	9 252	12. 55	103
英格兰	1 960	15. 43	69
德国	1 647	10. 93	52
意大利	1 573	11. 86	50
加拿大	1 342	14. 45	52
法国	1 266	12. 36	48
日本	1 023	8. 98	34
中国	943	7. 17	28
荷兰	827	17. 02	49
澳大利亚	734	16. 91	37

发表论文数量居前 10 位的国家均参与了 EBCTCG（早期乳腺癌多试验协作组）项目关于辅助化疗、内分泌治疗对乳腺癌复发及 15 年生存研究^[6]。至 2010 年 1 月 20 日本文采集数据时止，SCI 数据库中显示该文（“化疗与内分泌治疗早期乳腺癌复发与 15 年生存分析：随机对照试验”）被引用频次达 1 404 次，是乳腺癌相关领域 2005 – 2009 年内发表论文中被引用频次最高的 1 篇，该论文研究主题应归纳为“综合治疗”。与国际上的顶尖国家相比，中国的论文质量还有待提高，h 指数低于发文数量排在其后的西班牙、瑞士、瑞典、比利时、挪威、芬兰、丹麦、奥地利等国，排名为第 21 位。

3.2 高影响力机构及其被引频次最多的论文主题分布

经过分析，乳腺癌相关研究发文量最多的 10 个机构 2005 – 2009 年共计发表论文 4 931 篇，占文献总量的 19. 47%。其论文篇数、篇均被引次数、h 指数、单篇被引次数指标以及各机构被引次数最高的论文主题分布情况，见表 3。从论文篇数和 h 指数两个指标来看，美国德克萨斯大学、哈佛大学和美国国立癌症研究所排在前 3 位，是乳腺癌研究领域的高影响力机构。各个机构的单篇最高被引次数都在 200 次以上。

表 3 发表乳腺癌相关论文数量居前 10 位的机构及各机构被引频次最高的论文和主题分布

机构	论文量 (篇)	篇均被 引频次	h 指数	单篇最高 被引频次	被引频次最高的 论文题名	被引频次最高的 论文主题
德克萨斯大学	1 019	15. 36	53	1 404	化疗与内分泌治疗早期乳腺癌复发与 15 年生 存分析: 随机对照试验	综合治疗
哈佛大学	617	18. 12	48	523	来曲唑与三苯氧胺治疗绝经后早期乳腺癌比较	内分泌治疗
国立癌症研究所	535	20. 56	41	1 404	来曲唑与三苯氧胺治疗绝经后早期乳腺癌比较	内分泌治疗
多伦多大学	372	12. 66	34	295	胶片钼靶摄影与数字摄影用于乳腺癌筛查比较	同位素放射线诊断
斯隆凯特林癌症研究中心	351	23. 60	40	1 404	化疗与内分泌治疗早期乳腺癌复发与 15 年生 存分析: 随机对照试验	综合治疗
密西根大学	312	16. 74	37	285	美国临床肿瘤协会早期乳腺癌前哨淋巴结活检 指南	外科 - 前哨淋巴结
加州大学旧金山分校	301	20. 02	33	972	多基因预测三苯氧胺治疗的淋巴结阴性乳腺癌 复发	多基因表达谱预后
北卡罗来纳大学	295	18. 09	33	332	乳腺癌各分子亚型对术前化疗反应差异性研究	分子病理学
梅奥临床	263	27. 49	31	1 404	化疗与内分泌治疗早期乳腺癌复发与 15 年生 存分析: 随机对照试验	综合治疗
加州大学洛杉矶分校	244	19. 84	31	500	全基因组学乳腺癌新易感基因研究	易感基因

3. 3 高影响力作者及其被引频次最多的论文主题分布 (表 4, 表 5)

3. 4 被引频次最高的论文主题分布 (表 6)

表 4 发表乳腺癌相关论文数量居前 10 位的作者^[7-8]

作者	论文量 (篇)	篇均被 引频次	h 指数	单篇最高 被引频次
Hortobagyi, GN	123	24. 31	28	332
Goldhirsch, A	108	54. 07	23	1 404
Ellis, IO	84	17. 62	20	156
Altundag, K	77	1. 42	5	24
Zheng, W	76	8. 95	15	64
Easton, DF	75	35. 19	25	500
Viale, G	75	22. 75	18	426
Bernstein, L	75	9. 03	15	44
Hankinson, SE	71	32. 72	25	500
Kaufmann, M	71	56. 83	19	1 404

表 6 被引频次最高的 15 篇论文作者及主题分布

被引频次	期刊	作者	主题
1 404	<i>Lancet</i>	Abe O 等	综合治疗
1 103	<i>N Engl J Med</i>	Romond EH	靶向治疗
1 002	<i>N Engl J Med</i>	Piccart - Gebhart	靶向治疗
972	<i>N Engl J Med</i>	Paik S	基因表达谱预后
729	<i>Lancet</i>	Howell A	综合治疗
672	<i>Lancet</i>	Abe O	保守治疗
644	<i>Lancet</i>	Wang YX	基因表达谱预后
523	<i>Cell</i>	Orimo A	肿瘤发生
521	<i>Cancer Research</i>	Iorio MV	MicroRNA
515	<i>New Engl J Med</i>	Thurlimann B	内分泌治疗
500	<i>Nature</i>	Easton DF	易感基因
494	<i>New Engl J Med</i>	Geyer CE	靶向治疗
464	<i>J Clin Oncol</i>	Miller KD	靶向治疗
455	<i>Nature</i>	Minn AJ	肿瘤发生
419	<i>N Engl J Med</i>	Miller K	靶向治疗

表 5 单篇最高被引频次高于 100 的作者
被引频次最高的论文主题分布

作者	机构	被引频次最高 的论文主题
Hortobagyi, GN	MD 安德森癌症中心 (美国)	分子病理学
Goldhirsch, A	欧洲肿瘤研究所 (意大利)	综合治疗
Ellis, IO	诺丁汉大学 (英国)	三阴乳腺癌
Easton, DF	剑桥大学 (英国)	易感基因
Viale, G	欧洲肿瘤研究所 (意大利)	综合治疗
Hankinson, SE	范德堡大学 (美国)	易感基因
Kaufmann, M	法兰克福大学 (德国)	综合治疗

3. 5 乳腺癌研究热点确定

对乳腺癌相关领域高影响力国家、机构、作者 (单篇最高被引频次 100 次以上) 发表的被引频次最高的论文以及该领域 2005 - 2009 年被引频次最高的 15 篇论文的主题分布进行统计, 去除交叉重复, 共统计出 20 篇论文, 分布于 11 个主题, 将此 11 个主题确定为 2005 - 2009 年乳腺癌研究热点。见表 7。

表 7 2005 – 2009 年乳腺癌研究热点

主题	论文量 (篇)	主题	论文量 (篇)
靶向治疗	5	同位素放射线诊断	1
分子病理学与基因	3	易感基因	1
表达谱预后		保守治疗	1
肿瘤发生	2	外科学 – 前哨淋巴结	1
内分泌治疗	2	三阴乳腺癌	1
综合治疗	2	合计	20
MicroRNA	1		

4 讨论

4.1 内容分析法应用于医学学科热点研究的可行性

本文将普遍应用于社会科学研究的 内容分析法应用于医学学科热点分析，通过研究乳腺癌热点，经类目主题统计，共确定了 11 个热点研究主题。对各主题的高被引论文逐篇分析，结合国内外最新综述性文章，研究各主题发展的现状和趋势，发现这些热点研究方向与乳腺癌治疗模式从“可以耐受的最大治疗”转变到“最小有效治疗”的理性手术方式上来，坚持规范化治疗并重视个体化治疗的原则相一致，也与力求实现乳腺癌早期预防、早期诊断和早期治疗的目标相一致，是近年来乳腺癌研究的热点主题，研究结果具有可信性。《中图法》医学类目尤其是肿瘤学类目与社科学相比，体系完善，是进行规范化热点研究的重要基础。

4.2 内容分析法进行医学学科热点研究的局限性

医学文献产出量大，而内容分析法涉及到大量的文献阅读，工作量大，因此在学科选择上不宜过大。本文选择肿瘤学子学科进行研究，结合学科高影响力的国家、机构、作者的高被引文献可适当弥补文献样本量不足的缺陷。另外，选择被引频次高的文献来作为研究样本，进行内容分析，对于热点研究有一定的局限性。如某主题领域发表文章较多，但单篇文章不能达到较高的被引频次，则在小

样本的文献研究中不能成为研究热点，在未来的研究中，应辅以发文量进行分析^[9-10]。

本篇论文研究过程中，对各篇高被引文献进行了内容分析，基本把握了学科热点的发展现状。将内容分析法用于医学学科热点研究，是与临床和科研一线紧密相连的基层图书情报工作者的一次尝试，有待在日后的工作中逐步改进。

参考文献

1 王知津. 网络计量法与内容分析法比较研究 [J]. 图书馆研究, 2006, (6): 2 – 5.

2 张宏梁, 田玲, 张黎黎. 利用文献计量学研究学科热点初探 [J]. 医学信息学杂志, 2008, 29 (11): 11 – 15.

3 中国图书馆分类法编辑委员会. 中国图书馆分类法 [M]. 第 4 版. 北京: 北京图书馆出版社, 1999: 470 – 474.

4 蒋新. h 指数的数学解读和 h 指数数列 [J]. 科学学研究, 2009, 27 (2): 187 – 189.

5 邱均平. 发文量和 h 指数结合的高影响力作者评选方法研究——以图书情报学为例的实证分析 [J]. 图书馆论坛, 2008, 28 (6): 44 – 49.

6 Abe O. Effects of Chemotherapy and Hormonal Therapy for Early Breast Cancer on Recurrence and 15 – year Survival: an overview of the randomised trials [J]. Lancet, 2005, 365 (9472): 1687 – 1717.

7 Rouzier R. Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy [J]. Clinical Cancer Research, 2004, 11 (16): 5678 – 5685.

8 Ayers M. Gene Expression Profiles Predict Complete Pathologic Response to Neoadjuvant Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide Chemotherapy in Breast Cancer [J]. Journal of Clinical Oncology, 2004, 22 (12): 2284 – 2293.

9 崔雷. 学科主题演变的深度挖掘分析——以普通外科学为例 [J]. 医学信息学杂志, 2009, 30 (8): 5 – 10.

10 李海燕. 近十年国外对中草药研究的关注点——基于高频主题词的共现聚类分析 [J]. 情报学报, 2009, 28 (3): 395 – 400.