

· 数据挖掘研究 ·

基于关联规则及复杂网络分析古医籍中乳腺癌的用药规律

陈文莉¹, 李文林^{2*}, 史勤¹, 陈红宇¹, 周娴芳¹, 丁明明¹, 刘立伟¹, 杨丽丽², 杨涛³

(1. 上海市宝山区中西医结合医院肿瘤科, 上海 201999; 2. 南京中医药大学图书馆, 江苏 南京 210023; 3. 南京中医药大学信息技术学院, 江苏 南京 210023)

摘要:目的 探析中医古籍文献中乳腺癌相关方剂中蕴含的乳腺癌治疗用药规律。方法 采用 XMinerV2.4 进行 Fp-Growth 算法建模进行方药数据关联分析, 并采用加权有向图形式表示药物之间的配伍关系; 采用 LiquoriceV3.0 进行 Complex Network 建模进行方药网络分析, 对网络节点进行社团聚集分析, 利用社团结构表示药物配伍关系。结果 共收集内服方剂 136 例, 外用方剂 41 例。内服方核心药物是莪术、白术、当归, 莪术的配伍有: 三棱-莪术、莪术-淫羊藿、莪术-鹿角片、莪术-肉苁蓉、莪术-枸杞子。当归的配伍有: 当归-人参、当归-防风、当归-熟地; 白术的配伍有: 白术-石见穿、白术-南沙参和白术-枸杞子。外用药物中枯矾、白芷、僵蚕-羌活等有关联性, 同时以僵蚕为主, 形成了僵蚕-羌活、僵蚕-(川乌、羌活)、僵蚕-香附等不同药对或药组。通过复杂网络分析可以看出, 内服方中的核心药物群有益气温阳活血药: 党参、白术、黄芪、茯苓、淫羊藿、肉苁蓉、延胡索、莪术、山茱萸; 养阴解毒药: 南沙参、枸杞子、石见穿、露蜂房; 行气活血药: 当归、郁金、香附; 化痰活血药: 三棱、桃仁、鹿角片、丹参、延胡索、海藻。核心药物之外的配伍分析提示有昆布-牡蛎、白芥子-鹿角胶、枳壳-厚朴、玄参-白蔹等药对。外用方中的核心药物群为活血通阳解毒药: 当归、白蔹、没药、葱白; 杏仁、蓖麻仁; 红花-乳香。核心药物之外的配伍分析提示有活血温里止痛药: 当归-川芎-草乌; 祛风药: 僵蚕-羌活; 化痰行气药: 枳壳-香附-陈皮-半夏等药对。结论 运用关联规则与复杂网络的方法分析古籍文献中乳腺癌的用药规律, 可为探讨古代该病的病机特点及临床治疗提供参考, 开拓当代临床医生的治疗思路, 结合分析结果, 可将其中的典型药对尤其是当代临床少见的药对, 在乳腺癌复发转移而常规治疗乏效时使用, 为乳腺癌的现代临床研究提供思路与方向。

关键词: 古籍; 乳腺癌; 数据挖掘; 用药特色; 配伍规律

中图分类号: R249.1 文献标志码: A 文章编号: 1672-0482(2018)06-0625-05
DOI: 10.14148/j.issn.1672-0482.2018.0625

Analysis of Drug Use Rule of Breast Cancer in Ancient Literature Based on Association Rules and Complex Network

CHEN Wen-li¹, LI Wen-lin^{2*}, SHI Qin¹, CHEN Hong-yu¹, ZHOU Xian-fang¹, DING Ming-ming¹, LIU Li-wei¹, YANG Li-li², YANG Tao³

(1. Department of Oncology, Shanghai Baoshan District Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai, 201999, China; 2. Nanjing University of Chinese Medicine Library, Nanjing, 210023, China; 3. School of Information Technology, Nanjing University of Chinese Medicine, Nanjing, 210023, China)

ABSTRACT: OBJECTIVE To explore the drug use rule of breast cancer in ancient medical literature on the basis of association rules and complex network. **METHODS** XMinerV2.4 was used to mold Fp-Growth algorithm for correlation analysis of prescription data, and weighted directed graph was used to represent the compatibility relationship between drugs. LiquoriceV3.0 was used to model the complex Network for network analysis of drugs, and the network nodes were analyzed for community aggregation, and the community structure was used to express the drug compatibility relationship. **RESULTS** 136 cases of oral prescriptions and 41 cases of external prescriptions were collected. The core drugs of the oral administration are e zhu (Rhizoma Curcumae), bai zhu (angelica) and dang gui (Radix Angelicae Sensis). The compatibility of e zhu (Rhizoma Curcumae) includes san leng (Rhizoma Sparganii) - e zhu (Rhizoma Curcumae), e zhu (Rhizoma Curcumae) - yin yang huo (Herba Epimedii), e zhu (Rhizoma Curcumae) - lu jiao pian (Cornu Cervi Tablets), e zhu (Rhizoma Curcumae) - rou cong rong (Cistanche Herba), and e zhu (Rhizoma Curcumae) - gou qi zi (Fructus Lycii). The compatibility of dang gui (Radix Angelicae Sensis) is:

收稿日期: 2018-02-26

基金项目: 上海市优秀青年中医临床人才培养计划(ZYSNXD011-RC-XLXX-20110001); 上海市宝山区科委项目(16-E-15); 上海市宝山区中西医结合医院项目(201504); 上海市宝山区中西医结合医院自然培育项目(GZRPYJJ-201704)

作者简介: 陈文莉(1971-), 女, 博士, 副主任医师。*通信作者: 李文林, 女, 研究馆员, 主要从事中医药数据分析的研究, E-mail: njliwenlin@126.com

dang gui (Radix Angelicae Sinesis) - ren shen (Ginseng), dang gui (Radix Angelicae Sinesis) - fangfeng (Divaricate Saposhnikovia), dang gui (Radix Angelicae Sinesis) - shu di (Rehmannia Glutinosa); The compatibility of bai zhu (Angelica) is: bai zhu (Angelica) - shijianchuan, bai zhu (Angelica) - nan sha shen (Glehnia Root) and bai zhu (Angelica) - gou qi zi (Fructus Lycii). The correlation between ku fan (Dried Alum) and bai zhi (Angelica Dahurica) is the highest among the external medicines, and at the same time, jiang can (Bombyx Batryticatus) is the main medicine, forming different medicine pairs or groups such as jiang can (Bombyx Batryticatus) - qiang huo (Notopterygium Root), jiang can (Bombyx Batryticatus) - (chuan wu (monkshood), qiang huo (Notopterygium Root)), jiang can (Bombyx Batryticatus) - xiang fu (Nutgrass Galingale Rhizome), etc. Through complex network analysis, it can be seen that the core drug group in the oral prescription is beneficial to warming yang and activating blood: dang shen (Radix Codonopsis), bai zhu (Rhizoma Atractylodis Macrocephalae), huang qi (Radix Astragali), fu ling (Poria), yin yang huo (Herba Epimedii), rou cong rong (Herba Cistanches), yan hu suo (Rhizoma Corydalis), e zhu (Rhizoma Curcuma), shan zhu yu (Fructus Corni); Nourishing yin and resolve toxin: nan sha shen (Radix Adenophorae), gou qi zi (Fructus Lycii), shi jian chuan (Radix Clematidis), and lu feng fang (Nidus Vespa); Moving qi and activating blood herbs: dang gui (angelica), yu jin (Curcuma Aromatica) and xiang fu (Nutgrass Galingale Rhizome); Dissolving phlegm and quickening blood herbs: san leng (Sparganium), tao ren (Peach Kernel), lu jiao pian (Cornu Cervi Slice), dan shen (Salvia Miltiorrhiza), yan hu suo (Rhizoma Corydalis) and hai zao (Seaweed). The compatibility analysis outside the core drugs suggested that there were drug pairs such as kun bu (Kelp) - mu li (Oyster), bai jie zi (Mustard Seed) - lu jiao jiao (Antler Glue), zhi qiao (Bitter Orange) - hou po (Magnolia Bark), xuan shen (Figwort) - bai lian (Ampelopsis), etc. The core medicine in the external prescription is the antidote for promoting blood circulation and removing yang: dang gui (Radix Angelicae Sinensis), bai lian (Radix Ampelopsis), mo yao (Myrrh), ru xiang (Olibanum), cong bai (Bulbus Allii Fistulosi), xing ren (Semen Armeniacae Amarum), bi ma ren (Castor Bean) and hong hua (Flos Carthami). Analysis of compatibility outside the core drugs suggests that there are pain relievers for promoting blood circulation and warming the interior: dang gui (Radix Angelicae Sinensis), chuan xiong (Rhizoma Chuanxiong), cao wu (Radix Aconiti Kusnezoffii); Dispersing wind medicine: jiang can (Bombyx Batryticatus), qiang huo (Rhizoma et Radix Notopterygii); Resolving Phlegm and moving qi herbs: zhi qiao (Fructus Aurantii), xiang fu (Rhizoma Cyperi), chen pi (Pericarpium Citri Reticulatae), ban xia (Rhizoma Pinelliae), etc. **CONCLUSION** Using association rules and complex network methods to analyze the drug use rules of breast cancer in ancient literature can provide a reference for exploring the pathogenesis and clinical treatment of the disease in ancient times, expand the treatment thinking of contemporary clinicians, and combine the analysis results to use the typical drug pairs, especially those rare in contemporary clinical practice, when the conventional treatment for breast cancer recurrence and metastasis is ineffective, thus providing ideas and directions for modern clinical research of breast cancer.

KEY WORDS: ancient literature; breast cancer; data mining; characteristic of drug use; compatibility rule

乳腺癌是严重威胁女性健康的恶性肿瘤,占女性肿瘤发病的首位,女性癌症死亡第 6 位^[1]。我国古代没有乳腺癌之病名,但中医文献中“乳痞”“乳岩”“乳核”“石痈”“石榴翻花发”等症状描述与乳腺癌症状类似,相关学者已从学术源流角度进行了梳理与考证^[2-4],因此基于古文献中关于乳腺癌病案中积累的大量方药进行分析,可为当代中医临床对本病的预防与治疗提供重要的借鉴与参考。

关联规则与复杂网络是以网络化建模形式研究复杂现象的一种分析方法,是一种描述组成复杂系统各元素间关系的一种表达形式,在中医药研究领域已有相对成熟的应用^[5]。为此,本研究以南京中医药大学图书馆古籍文献数据库中可采集到的古代乳腺癌相关病案方剂为研究对象,运用关联规则与复杂网络进行分析,得出中医药治疗本病的核心药物及配伍规律,以期为临床提供依据。

1 资料与方法

1.1 资料来源

相关方剂来源于南京中医药大学图书馆中医古

籍文献数据库中收录的《外科全生集》《饲鹤亭集方》《医方简义》《外科正宗》《医宗金鉴》《疡医大全》《圣济总录》《喉科心法》《医门补要》《医林纂要》《外科大成》《外科集腋》《千金方》《集验良方》《理瀹》《青囊秘诀》《医门八法》《顾氏医径》《惠直堂方》《古方汇精》《揣摩有得集》《扁鹊心书·神方》《竹林女科》《医事启源》《疡科选粹》《疡科心得集·方汇补遗》《仙拈集》《玉机微义》《外台秘要》等医籍,版本以明清年代为主。另有《医部全录》《药庵医学丛书·论医集》《奇方类编》等民国医籍。

1.2 纳入标准

相关古籍中收录的方剂“功效”或“主治”中含“乳岩”“乳石痈”“奶岩”“癆岩”“妒乳”“石奶”“翻花奶”“乳栗”“乳痞”“乳衄”中的任意一项,且症状描述与当代乳腺癌症状一致者。

1.3 排除标准

①方剂功效或主治病症描述与现代乳腺癌症状不符或无关者;②只有方剂名称而无具体方药组成者。

1.4 数据规范与预处理

①对收集的方剂原始数据进行预处理,根据古籍中描述的使用方法将纳入方剂分为内服方剂和外用方剂;②参照《中药学》^[6]教材对药物名称予以规范化处理。

2 数据分析方法

采用 SPSS18.0 统计软件进行描述性统计;采用 XMinerV2.4 进行 Fp-Growth 算法建模并进行方药数据关联分析,采用加权有向图表示药物之间的配伍关系;采用 LiquoriceV3.0 进行 Complex Network 建模进行方药网络分析,对网络节点进行社团聚集分析,利用社团结构表示药物之间配伍关系。

3 结果

3.1 药物频次

共收集内服方剂 136 例,外用方剂 41 例。结果见表 1~2。

3.2 关联规则分析

置信度、支持度、提升度是关联规则的常用指标。置信度反映了项集在事务数据库中出现的频数;支持度反映了项集出现的条件概率;提升度反映了项集之间的相关性,提升度越大,项集之间的关联意义越大。提升度 >1 且越高表明正相关性越高,提升度 <1 且越低表明负相关性越高,提升度 $=1$ 表明没有相关性,即相互独立。提升度是比传统支持度、置信度更能反映关联规则中事务之间的相关性。因此,本文将提升度作为规则衡量的主要指标。

表 1 内服方前 30 味药物频次分布

药物	频次	药物	频次	药物	频次
当归	69	蒲公英	19	山慈菇	14
甘草	33	白芷	18	肉苁蓉	14
茯苓	30	莪术	18	生甘草	13
白芍	29	青皮	16	金银花	13
白术	25	乳香	16	没药	13
党参	24	山茱萸	16	生黄芪	13
黄芪	23	陈皮	16	连翘	12
川芎	22	川贝母	16	香附	12
夏枯草	21	淫羊藿	15	漏芦	12
柴胡	19	人参	14	麝香	11

设置最小支持度(minSup)、最小置信度(minConf)、最小提升度(minLift)阈值,基于 Fp-Growth 算法,从频繁项集中得到内服方药及外用方药的配伍关联关系表及可视化表征结果,结果见表 3~6 及图 1~2。

表 2 外用方前 30 味药物频次分布

药物	频次	药物	频次	药物	频次
冰片	6	川芎	4	麝香	4
当归	6	乌药	4	蟾酥	4
川乌	5	香附	4	轻粉	4
白芷	5	白芷	4	半夏	3
乳香	5	白及	4	陈皮	3
没药	5	生姜	4	青皮	3
血竭	5	葱白	4	羌活	3
儿茶	5	槐枝	4	枳壳	3
香油	5	穿山甲	4	僵蚕	3
硃砂	4	樟脑	4	白芥子	3

表 3 内服方两味药之间的关联关系

序号	规则	支持度	置信度	提升度
1	瓜蒌仁→漏芦	0.051	1	11.333
2	鹿角片→肉苁蓉	0.059	1	9.714
3	鹿角片→淫羊藿	0.059	1	9.067
4	枸杞子→莪术	0.059	1	7.556
5	鹿角片→莪术	0.059	1	7.556
6	枸杞子→党参	0.059	1	5.667
7	枸杞子→白术	0.059	1	5.440
8	枸杞子→茯苓	0.059	1	4.690
9	熟地→当归	0.044	1	2.030
10	泽兰→当归	0.044	1	2.030

注:设置最小支持度为 0.04,最小置信度为 0.9,最小提升度为 1。

表 4 内服方多味药之间的关联关系

序号	规则	支持度	置信度	提升度
1	延胡索,丹参→三棱,山慈菇	0.044	1	19.429
2	延胡索,三棱→丹参,山慈菇	0.044	1	19.429
3	延胡索,丹参→三棱	0.044	1	15.111
4	延胡索,山慈菇→三棱	0.044	1	15.111
5	延胡索,三棱→丹参	0.044	1	12.364
6	延胡索,山慈菇→丹参	0.044	1	12.364
7	瓜蒌仁,白芷→漏芦	0.044	1	11.333
8	延胡索,丹参→山慈菇	0.044	1	9.714
9	延胡索,三棱→山慈菇	0.044	1	9.714
10	延胡索,丹参,三棱→山慈菇	0.044	1	9.714

注:设置最小支持度为 0.04,最小置信度为 0.9,最小提升度为 1。

3.3 复杂网络分析

方剂配伍是一个复杂系统,而复杂网络是研究复杂系统的有效方法之一,它从图论角度揭示复杂系统的组织结构和演化规律。聚集系数是复杂网络的一项重要参数,它反映了网络中节点的聚集程度,能较好地用来度量药物之间的相互依赖情况。聚类系数越大,表示药物与药物之间的依赖关系越强。通过设定聚集系统,网络自动划分社群结构,同一社群的节点联系紧密,可以认为是常用的中药配伍。

本次研究中,内服方剂数量较多,研究中取聚类系数为 150,外用方剂数量相对较少,取聚类系数为 70,分别获得两类方剂药物之间的整体配伍关系图,见图 3~4。

表 5 外用方两味药关联关系

(minSup=0.05, minConf=0.8)

序号	规则	支持度	置信度	提升度
1	僵蚕→羌活	0.073	1	13.667
2	羌活→僵蚕	0.073	1	13.667
3	枯矾→白芷	0.073	1	10.250
4	僵蚕→香附	0.073	1	10.250
5	僵蚕→川乌	0.073	1	8.200

注:设置最小支持度为 0.05,最小置信度为 0.8,最小提升度为 1。

表 6 外用方多味药关联关系 (minSup=0.05, minConf=0.8)

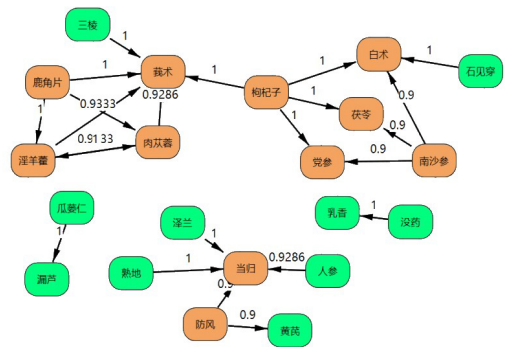
序号	规则	支持度	置信度	提升度
1	僵蚕→川乌,羌活	0.073	1	13.667
2	僵蚕,川乌→羌活	0.073	1	13.667
3	羌活→僵蚕,川乌	0.073	1	13.667
4	川乌,羌活→僵蚕	0.073	1	13.667
5	僵蚕→川乌,香附	0.073	1	13.667
6	川乌,香附→僵蚕	0.073	1	13.667
7	僵蚕→羌活,香附	0.073	1	13.667
8	羌活→僵蚕,香附	0.073	1	13.667
9	僵蚕,香附→羌活	0.073	1	13.667
10	羌活,香附→僵蚕	0.073	1	13.667
11	僵蚕→川乌,羌活,香附	0.073	1	13.667
12	僵蚕,川乌→香附	0.073	1	10.250
13	僵蚕,羌活→香附	0.073	1	10.250
14	僵蚕,羌活→川乌	0.073	1	8.200
15	僵蚕,香附→川乌	0.073	1	8.200

注:设置最小支持度为 0.05,最小置信度为 0.8,最小提升度为 1。

4 讨论

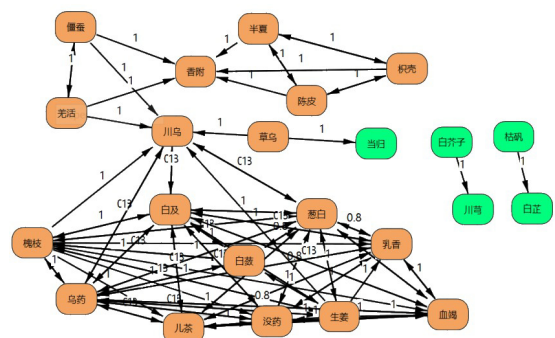
4.1 不同药对的配伍体现治疗原则的组合

关联规则分析反映了两药之间的配伍紧密程度,依赖关系网络分析可以发现核心药物,两者结合则可以发现核心药组。内服方核心药物是“莪术、白术、当归”。如莪术的配伍除了活血化瘀三棱-莪术药对外,还有莪术-淫羊藿、莪术-鹿角片、莪术-肉苁蓉等补阳活血药对和莪术-枸杞子活血滋阴药对。当归的配伍则有当归-人参补气养血药对、当归-防风养血祛风药对、当归-熟地补血药对;白术则有白术-石见穿补气活血药对、白术-南沙参和白术-枸杞子等补气养阴药对。



注:最小支持度为 0.04,最小置信度为 0.9,最小提升度为 1。

图 1 内服方配伍依赖关系网络



注:最小支持度为 0.05,最小置信度为 0.8,最小提升度为 1。

图 2 外用方配伍依赖关系网络

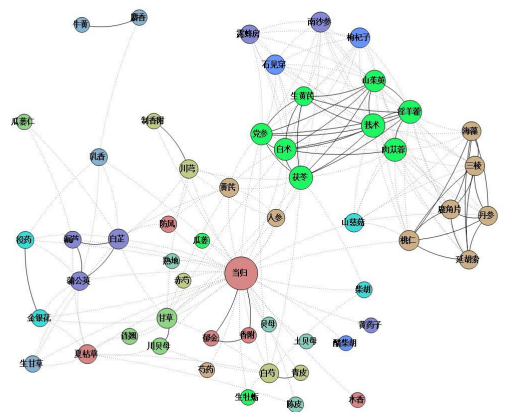


图 3 内服方核心药物网络关系图

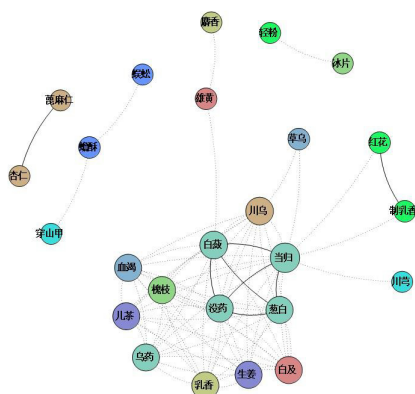


图4 外用方核心药物网络关系图

还有一些多药关联药对,如化痰活血类:山慈菇-(延胡索、丹参);山慈菇-(延胡索、丹参、三棱)。化痰解毒类:瓜蒌仁-漏芦。从临床来看,乳腺癌在肿块形成时以气滞血瘀为主要病机,但是后期会出现冲任不调,特别是晚期可能会发生肝、肺转移,胸腹水等痰浊水湿不化的情况,逐渐发展至脾肾阳虚,正虚毒炽状态。这些药对的出现与当代临床实际具有一致的特征。

外用药物中桔矾-白芷、僵蚕-羌活等有关联性,同时以僵蚕为主,形成了僵蚕-羌活、僵蚕-川乌-羌活、僵蚕-香附等不同药对或药组。从现代临床可见,乳腺癌在中晚期肿块局部扩散,会发生局部压迫疼痛或者骨转移导致的疼痛,治疗多以活血行气立法,故乳香-没药-儿茶-血竭的配伍较多。此外,本研究还提示古人擅长外用川乌止痛,配伍行气、解毒等药。图2揭示了外用药物之间的配伍频繁项集,其中关联度较高的药物有川乌-羌活、川乌-槐枝、川乌-乌药等。三棱-莪术在乳腺癌的治疗中适用广泛,活血法可能是治疗乳腺癌的常用治法,为临床使用提供参考。

4.2 内服方重视整体辨证,外用方着眼标实

通过复杂网络分析可以看出,内服方中的核心药物群有益气温阳活血药:党参、白术、黄芪、茯苓、淫羊藿、肉苁蓉、延胡索、莪术、山茱萸;养阴解毒药:南沙参、枸杞子、石见穿、露蜂房;行气活血药:当归、

郁金、香附;化痰活血药:三棱、桃仁、鹿角片、丹参、延胡索、海藻。核心药物之外的配伍分析提示有昆布-牡蛎、白芥子-鹿角胶、枳壳-厚朴、玄参-白蔹等药对。外用方中的核心药物群为活血通阳解毒药:当归、白蔹、没药、葱白;杏仁、蓖麻仁;红花、乳香。核心药物之外的配伍分析提示有活血温里止痛药:当归-川芎-草乌;祛风药:僵蚕-羌活;化痰行气药:枳壳-香附-陈皮-半夏等药对。总之内服方组方更加复杂,根据患者的体质、病机及兼证的不同可以选择不同配伍;外用方中治法相对单纯,治标为先。

现代临床治疗乳腺癌多为术后放、化疗后或者靶向治疗后的患者,局部肿瘤负荷大大减少,正气偏虚,以扶正抑瘤为主要治则,拟益气养阴、调节冲任、疏肝理气和胃、补益肝肾等法,辅以清热解毒、软坚散结、活血化瘀、清热利湿等治法。但是对于晚期复发的患者或者无法手术的晚期患者,阳气不足,癌毒炽盛,仍可以借鉴古代的治疗方法方药处理。

运用关联规则与复杂网络的方法分析古籍文献中乳腺癌的用药规律,可为探讨古代该病的病机特点及临床治疗提供参考,开拓当代临床医生的治疗思路,结合分析结果,可将其中的典型药对尤其是当代临床少见的药对,在乳腺癌复发转移而常规治疗乏效时使用,为乳腺癌的现代临床研究提供思路与方向。

参考文献:

- [1] 陈万青,张思维,曾红梅等.中国2010年恶性肿瘤发病与死亡[J].中国肿瘤,2014,23(1):1-10.
- [2] 张文娟.乳岩中医证治的文献研究[D].济南:山东中医药大学,2008.
- [3] 毕丽娟,司富春,王峰.古代治疗乳岩的方药分析[J].时珍国医国药,2006,17(4):656-657.
- [4] 葛洪.肘后备急方[M].北京:人民卫生出版社,1983:137-155.
- [5] 李文林,赵国平,陆建峰,李欣.关联规则在名医临证经验分析挖掘中的应用[J].南京中医药大学学报(自然科学版),2008,24(1):21-24.
- [6] 凌一揆,颜正华.中药学[M].上海:上海科技出版社,2017:5.

(编辑:叶亮)