

基于关联规则与熵聚类的清热类中成药组方规律研究

吴嘉瑞 金燕萍 张晓朦 张 冰 盛晓光

(北京中医药大学,北京,100102)

摘要 目的:探讨常用清热类中成药组方规律。方法:收录《新编国家中成药》中清热类中成药处方,采用关联规则 Apriori 算法和复杂系统熵聚类等方法,确定处方中药物的使用频次及药物之间的关联规则等。结果:高频次药物包括甘草、黄芩、冰片、金银花、大黄、连翘等;高频次药物组合包括“甘草、黄芩”“甘草、桔梗”“连翘、金银花”等;置信度较高的关联规则包括“牛黄 - > 冰片”“桔梗 - > 甘草”“朱砂 - > 冰片”“栀子 - > 黄芩”等。结论:处方用药中除常见的清热类中药外,尚包括具有清热作用的部分开窍药、安神药、泻下药及其他类药物。

关键词 清热药;关联规则;熵聚类

Analysis on the Composition Rules of Chinese Patent Medicine with Heat-clearing Based on Association Rules and Clustering Algorithm

Wu Jiarui, Jin Yanping, Zhang Xiaomeng, Zhang Bing, Sheng Xiaoguang

(Beijing University of Chinese Medicine, Beijing 100102)

Abstract Objective: To investigate composition rules of with heat-clearing Chinese patent medicine. **Methods:** The prescriptions of Chinese patent medicine with heat-clearing in “The new national Medicine” were collected to build a database of the methods of association rules with apriori algorithm and complex system entropy cluster were used to achieve the frequency of medicines and association rules between drugs. **Results:** Data-mining results indicated that in the prescriptions of Chinese patent drugs with heat-clearing, the most frequently used drugs were Radix Glycyrrhizae, Radix Scutellariae, Borneol, Flos Lonicerae Japonicae, Radix et Rhizoma Rhei, and Fructus Forsythiae. The most frequently used drug combinations were “Radix Glycyrrhizae with Radix Scutellariae”, “Radix Glycyrrhizae with Radix Platycodi”, and “Fructus Forsythiae with Flos Lonicerae Japonicae”, etc. Medicine with a high degree confidence coefficient of association rules include “Bezoar - > Borneol”, “Radix Platycodi - > Radix Glycyrrhizae”, “Cinnabisi - > borneol”, and “Fructus Gardeniae - > Radix Scutellariae”. **Conclusion:** In the prescriptions of Chinese patent drugs, there includes not only the drugs with heat-clearing property, but also that with resuscitation, calming, lapactic and other properties.

Key Words Chinese patent drugs; Heat-clearing; Association Rules; Entropy Clustering

中图分类号:R289.9;R311 文献标识码:A doi:10.3969/j.issn.1673-7202.2015.03.034

清热药是指以清解里热,治疗里热证为主的药物。根据药物来源及应用特点不同,清热类中药分为清热泻火、清热燥湿、清热凉血、清热解毒和清虚热药五类。清热类中成药种类繁多,其处方中除包含常见的清热泻火、清热解毒等五类清热药外,亦常配伍其他类别药物,配伍组合规律值得深入探讨。本研究收集《新编国家中成药》中的清热类中成药处方,在构建数据库的基础上,应用关联规则和熵聚类算法进行处方用药规律研究^[1-3]。

1 资料与方法

1.1 处方来源与筛选

本研究以《新编国家中

药》(第二版)^[4]为来源,共筛选出清热药处方1202首。

1.2 分析软件 “中医传承辅助系统(V2.0)”软件^[5],中国中医科学院中药研究所提供。该软件集“数据录入—数据管理—数据查询—数据分析—网络可视化展示”为一体,可以有效实现病案、疾病、症候、中药、处方、四诊等信息的管理、查询、综合分析等功能,可用于处方组方规律分析及新药处方发现等领域^[6-9]。

1.3 处方的录入与核对 将筛选后的处方录入“中医传承辅助系统(V2.0)”,录入完成后,由双人负责

基金项目:国家科技支撑计划课题(编号:2007BAI10B01);北京市中医药科技发展基金课题(编号:JJ-2010-70);北京中医药大学科研创新团队项目资助(编号:2011-CXTD-14);北京中医药大学“重点学科”开放课题(编号:2013-ZDXKKF-19)

作者简介:吴嘉瑞,副教授,博士,硕士研究生导师,研究方向:临床中医学, E-mail: exogamy@163.com

通信作者:张冰(1959.8—),女,博士,教授,博士研究生导师,研究方向:临床中医学, E-mail: zhangbing@263.net

数据的审核,以确保数据的准确性。

1.4 数据分析

1.4.1 频次统计分析 将清热处方中每味药的出现频次从大到小排序,并将“频次统计”结果导出。

1.4.2 组方规律分析 经过预实验,确定在支持度(表示在所有药物中同时出现的次数)为 60、置信度 ≥ 0.4 条件下开展研究,按药物组合出现频次从大到小的顺序进行排序;利用软件中“规则分析”功能运算得关联规则。

1.4.3 新方分析 在相关度为 8,惩罚度为 4 的条件下,应用改进的互信息法和复杂系统熵聚类方法发现核心组合与新组方,并实现网络可视化展示。

2 结果

2.1 用药频次分析 将药物按使用频次从高到低进行排序,前三位分别是甘草、黄芩和冰片,频次前 30 位的药物情况见表 1。

表 1 处方中使用频次前 30 位的药物

序号	中药名称	频次	序号	中药名称	频次
1	甘草	324	16	朱砂	90
2	黄芩	255	17	防风	89
3	冰片	207	18	赤芍	89
4	金银花	177	19	当归	88
5	大黄	163	20	柴胡	88
6	连翘	158	21	穿心莲	88
7	黄连	145	22	蒲公英	85
8	桔梗	130	23	牛黄	83
9	栀子	125	24	茯苓	80
10	薄荷	113	25	黄芪	76
11	黄柏	110	26	人工牛黄	73
12	板蓝根	106	27	白芷	70
13	生地黄	103	28	白芍	69
14	石膏	95	29	人工麝香	67
15	玄参	91	30	陈皮	66

表 2 处方中高频次药物组合

序号	药物组合	频次	序号	药物组合	频次
1	甘草,黄芩	99	10	金银花,黄芩	69
2	甘草,桔梗	91	11	黄芩,黄连	67
3	连翘,金银花	87	11	牛黄,冰片	66
4	甘草,连翘	84	13	甘草,冰片	66
5	栀子,黄芩	81	14	冰片,黄连	66
6	甘草,金银花	76	15	朱砂,冰片	63
7	连翘,黄芩	75	16	连翘,桔梗	61
8	大黄,黄芩	72	17	甘草,黄连	60
9	甘草,薄荷	71	18	大黄,甘草	60

2.2 基于关联规则分析的组方规律分析 按照药物组合出现频次由高到低排序,前 3 位分别是“甘草、黄芩”“甘草、桔梗”“连翘、金银花”。出现频次 60 次以上的药物组合见表 2。在支持度为 60,置信

度 ≥ 0.4 条件下,分析得药物关联规则,结果见表 3,关联规则网络图见图 1。

表 3 处方中药物组合关联规则(置信度 ≥ 0.4)

序号	规则	置信度	序号	规则	置信度
1	牛黄 -> 冰片	0.7951	9	连翘 -> 黄芩	0.4746
2	桔梗 -> 甘草	0.7000	10	桔梗 -> 连翘	0.4692
3	朱砂 -> 冰片	0.7000	11	黄连 -> 黄芩	0.4620
4	栀子 -> 黄芩	0.6480	12	黄连 -> 冰片	0.4551
5	薄荷 -> 甘草	0.6283	13	大黄 -> 黄芩	0.4417
6	连翘 -> 金银花	0.5506	14	金银花 -> 甘草	0.4293
7	连翘 -> 甘草	0.5316	15	黄连 -> 甘草	0.4137
8	金银花 -> 连翘	0.4915			

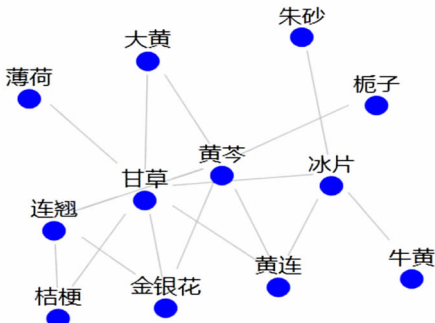


图 1 支持度为 60,置信度为 0.4 条件下的关联规则网络展示图

2.3 基于熵聚类的组方规律分析

2.3.1 基于改进的互信息法的药物间关联度分析

依据处方数量,结合经验判断和不同参数提取数据的预读,设置相关度为 8,惩罚度为 4,进行聚类分析,得到处方中两两药物间的关联度,将关联系数 0.015 以上的药对列表。见表 4。

表 4 基于改进的互信息法的药物间关联度分析

序号	药对	关联系数	序号	药对	关联系数
1	朱砂,全蝎	0.0343	16	桔梗,前胡	0.0172
2	朱砂,珍珠	0.0263	17	全蝎,人工麝香	0.0171
3	朱砂,天麻	0.0257	18	冰片,僵蚕	0.0171
4	朱砂,水牛角	0.0251	19	桔梗,黄芩	0.017
5	天竺黄,牛黄	0.0211	20	天麻,牛黄	0.0167
6	冰片,胆南星	0.0204	21	琥珀,冰片	0.0164
7	天麻,白附子	0.0202	22	胆南星,人工麝香	0.0161
8	朱砂,黄连	0.0196	23	珍珠,硼砂	0.0161
9	冰片,水牛角	0.0196	24	朱砂,羚羊角	0.0157
10	连翘,防风	0.0191	25	人工麝香,钩藤	0.0157
11	冰片,儿茶	0.0189	26	薄荷,菊花	0.0155
12	天竺黄,冰片	0.0183	27	牛黄,僵蚕	0.0153
13	全蝎,白附子	0.0182	28	薄荷,淡竹叶	0.0151
14	胆南星,钩藤	0.0176	29	琥珀,钩藤	0.015
15	全蝎,冰片	0.0173	30	冰片,钩藤	0.015

2.3.2 基于复杂系统熵聚类的药物核心组合与新处方分析 以改进的互信息法的药物间关联度分析结果为基础,按照相关度与惩罚度约束,基于复杂系

统熵聚类,演化出 3~4 味药物核心组合,具体见表 5。在核心组合提取的基础上,运用无监督熵层次聚类算法,得到 34 个新处方,具体见表 6。

表 5 基于复杂系统熵聚类的药物核心组合		
序号	核心组合	
1	淫羊藿,王不留行,菟丝子	菟丝子,山药,沙苑子
2	生地黄,麦冬,玄参	生地黄,赤芍,当归
3	生地黄,赤芍,玄参	赤芍,天花粉,玄参
4	蛇床子,白矾,苦参	蛇床子,黄柏,苦参,地肤子
5	半边莲,半枝莲,重楼	半枝莲,重楼,鳖甲
6	板蓝根,茵陈,柴胡	茵陈,金钱草,柴胡
7	桑白皮,苦杏仁,前胡	麻黄,苦杏仁,前胡
8	乳香,莪术,没药	乳香,没药,木鳖子
9	丁香,升麻,滑石	瞿麦,木通,滑石
10	甘草,苦杏仁,桔梗	甘草,薄荷,荆芥,桔梗,连翘
11	甘草,荆芥,淡竹叶	甘草,薄荷,荆芥,防风
12	三棱,威灵仙,乌药	独活,威灵仙,骨碎补
13	白术,茯苓,党参	泽泻,猪苓,茯苓
14	香墨,熊胆粉,牛胆汁	熊胆粉,珍珠母,炉甘石
15	王不留行,海金沙,金钱草	延胡索,海金沙,金钱草,鸡内金
16	三七,蒲黄,海螵蛸	三七,海螵蛸,白及
17	布渣叶,水翁花,香薷	水翁花,香薷,木棉花
18	火炭母,金樱根,木蝴蝶	金樱根,木蝴蝶,五指柑
19	黄连,雄黄,冰片,牛黄	雄黄,珍珠,冰片,牛黄,人工麝香
20	樟脑,轻粉,甲硝唑	樟脑,轻粉,地塞米松
21	地耳草,垂盆草,黄花倒水莲	地耳草,垂盆草,徐长卿
22	川芎,白芷,防风	细辛,辛夷,苍耳子,白芷
23	广金钱草,金沙藤,鸭脚木皮	布渣叶,广金钱草,金沙藤,木蝴蝶
24	六神曲,山楂,麦芽	山楂,麦芽,稻芽
25	香薷,槟榔,厚朴	枳实,槟榔,厚朴
26	牛蒡子,薄荷,荆芥,桔梗	牛蒡子,薄荷,荆芥,桔梗,连翘
27	枸杞子,黄芪,丹参	白花蛇舌草,黄芪,丹参
28	白矾,苦木,硼酸	珍珠,硼酸,冰片
29	泽泻,猪苓,车前子	泽泻,川木通,车前子
30	金银花,玄参,连翘	牛蒡子,淡豆豉,金银花,淡竹叶
31	儿茶,硼砂,青黛	硼砂,青黛,玄明粉
32	苦参,地肤子,何首乌	白鲜皮,花椒,苦参,何首乌

3 讨论

本研究应用关联规则和熵聚类算法系统分析了《新编国家中成药》中清热类中成药的用药规律。以下结合研究结果,对方剂中高频次药物进行分析。甘草是出现频率最高的单味中药,味甘性平,归心脾肺胃经。功能补脾益气,祛痰止咳,缓急止痛,清热解毒,且在处方中,甘草常发挥调和药性的作用,可降低方中某些药物(如大黄)的峻烈之性^[10-11],故其虽非典型清热药物,但却在处方中频次最高。黄芩出现频次仅次于甘草,其味苦性寒,归肺胆脾胃大小肠经,具有清热燥湿、泻火解毒、止血、安胎的功效,尤长于清中上焦湿热,可用于治疗湿温暑湿之证以及肺热壅遏所致咳嗽痰稠等。冰片味辛苦,微寒,归心脾肺经,能开窍醒神、清热止痛。金银花甘寒,有

清热解毒,疏散风热之功,为治一切内痈外痈之要药。连翘苦,微寒,清热解毒,消肿散结,疏散风热,主入心经,清心火,解疮毒的同时能消散痈肿结聚,为“疮家圣药”^[12]。

表 6 基于熵层次聚类的新处方	
序号	新处方
1	淫羊藿,王不留行,菟丝子,山药,沙苑子
2	生地黄,麦冬,玄参,赤芍,当归
3	生地黄,赤芍,玄参,天花粉
4	蛇床子,白矾,苦参,黄柏,地肤子
5	半边莲,半枝莲,重楼,鳖甲
6	板蓝根,茵陈,柴胡,金钱草
7	桑白皮,苦杏仁,前胡,麻黄
8	乳香,莪术,没药,木鳖子
9	丁香,升麻,滑石,瞿麦,木通
10	甘草,苦杏仁,桔梗,薄荷,荆芥,连翘
11	甘草,荆芥,淡竹叶,薄荷,防风
12	三棱,威灵仙,乌药,独活,骨碎补
13	白术,茯苓,党参,泽泻,猪苓
14	香墨,熊胆,牛胆汁,珍珠母,炉甘石
15	王不留行,海金沙,金钱草,延胡索,鸡内金
16	三七,蒲黄,海螵蛸,白及
17	儿茶,硼砂,青黛,玄明粉
18	苦参,地肤子,何首乌,白鲜皮,花椒
19	牛蒡子,薄荷,桔梗,荆芥,连翘
20	黄连,雄黄,冰片,牛黄,珍珠,人工麝香
21	地耳草,垂盆草,黄花倒水莲,徐长卿
22	川芎,白芷,防风,细辛,辛夷,苍耳子
23	广金钱草,金沙藤,鸭脚木皮,布渣叶,木蝴蝶
24	六神曲,山楂,麦芽,稻芽
25	香薷,槟榔,厚朴,枳实
26	金银花,玄参,连翘,牛蒡子,淡豆豉,淡竹叶
27	枸杞子,黄芪,丹参,白花蛇舌草
28	白矾,苦木,硼酸,珍珠,冰片
29	泽泻,猪苓,车前子,川木通

本研究成果的药物间关联规则有助于分析药物间的关联程度,如“牛黄 -> 冰片”(置信度为 0.795 1)含义为,处方中有牛黄时,有冰片的概率为 79.51%;“桔梗 -> 甘草”(置信度为 0.700 0)含义为,处方中有桔梗时,有甘草的概率为 70%;“朱砂 -> 冰片”(置信度为 0.700 0)含义为,处方中有朱砂时,有冰片的概率为 70%。又如关联规则网络图中共 12 味药物,其中除朱砂、金银花外,其余 10 味药均为名方牛黄上清丸^[13]中的药物,可见中成药处方中高关联度药物组合与中医经典名方的一致性,这也在一定程度上体现了中医理论的传承与延续。

再者,本研究应用复杂系统熵聚类方法得出的新处方不乏启迪意义,如序号 26 的新处方“金银花、玄参、连翘、牛蒡子、淡豆豉、淡竹叶”,其中金银花清热解毒,玄参清热凉血,连翘消肿散结,牛蒡子疏散

风热,淡豆豉宣发郁热,淡竹叶清热泻火,诸药合用既能清里热,又能解表邪,达到表里双解,祛邪扶正之功效。

综上,本研究应用数据挖掘方法对《新编国家中成药》中的清热类中成药组方规律进行了系统分析,研究所得结果如高频次药物、药物组合、高置信度关联规则等对于阐明中医清热类处方规律提供了参考,同时有助于清热类中药处方的优化与新药研发时的辅助设计。当然,应用关联规则和熵聚类所得的组方规则等还需进一步临床验证,数据挖掘方法亦有其局限性,需结合中医药理论与实践综合分析、评价^[14-16]。

参考文献

- [1]李瑞奇,白明,苗明三. 清热药的特点及现代研究[J]. 中医学报, 2013, 28(8): 1003-1005.
- [2]唐仕欢,陈建新,杨洪军,等. 证-熵-方—中医药研究的新领域[J]. 中国中医基础医学杂志, 2010, 16(4): 267-270.
- [3]唐仕欢,杨洪军. 中医组方用药规律研究进展述评[J]. 中国实验方剂学杂志, 2013, 19(5): 359-363.
- [4]宋民宪,杨明. 新编国家中成药[M]. 2 版. 北京: 人民卫生出版社, 2011.
- [5]卢朋,李健,唐仕欢,等. 中医传承辅助系统软件开发与应用[J]. 中国实验方剂学杂志, 2012, 18(9): 1-4.
- [6]杨洪军,赵亚丽,唐仕欢,等. 基于熵方法分析中风病处方中药物之间的关联度[J]. 中国中医基础医学杂志, 2005, 11(9): 706-

709.

- [7]唐仕欢,陈建新,杨洪军,等. 基于复杂系统熵聚类方法的中药新药处方发现研究思路[J]. 世界科学技术-中医药现代化, 2009, 11(2): 225-228.
- [8]吴嘉瑞,唐仕欢,郭位先,等. 基于数据挖掘的名老中医经验传承研究述评[J]. 中国中药杂志, 2014, 39(4): 614-617.
- [9]盛情,刘绍燕,庄曾渊,等. 基于中医传承辅助系统的李东垣眼病方剂用药规律研究[J]. 中国中医眼科杂志, 2014, 24(3): 177-180.
- [10]苗明三,张玉林,杨亚蕾,等. 基于中药药理作用的中药药性理论研究[J]. 时珍国医国药, 2009, 20(8): 2013-2015.
- [11]高学敏. 中药学[M]. 2 版. 北京: 中国中医药出版社, 2009: 432-433.
- [12]孟长海,王治英. 话说“疮家圣药”—连翘[J]. 环球中医药, 2008, 1(2): 19-19.
- [13]孙萍. 四种上清丸在临床中的辨证应用[J]. 中医临床研究, 2011, 3(18): 55-55.
- [14]李健,卢朋,唐仕欢,等. 基于中医传承辅助平台的治疗肺病方剂组方规律分析[J]. 中国实验方剂学杂志, 2012, 18(2): 254-254.
- [15]卢笑晖,单崎玮. 基于中医传承辅助平台分析卢尚岭教授治疗头风病用药经验[J]. 中国实验方剂学杂志, 2012, 18(9): 5-5.
- [16]吴嘉瑞,郭位先,张冰,等. 基于数据挖掘的国医大师颜正华含陈皮处方用药规律研究[J]. 中国中药杂志, 2014, 39(4): 618-622.

(2014-07-19 收稿 责任编辑: 张文婷)

(上接第 433 页)

有较强的导向作用,引导被评估者向目标靠近。因此,中医临床研究伦理质控评估指标就中医临床研究伦理质量控制的关键环节进行评估,对课题组提高临床研究伦理质量控制水平有明显的导向作用。同时,在试验早期阶段伦理质控的水平对于课题整体质量控制的影响更为重要。

参考文献

- [1]汪秀琴,熊宁宁,王思成. 中医药临床研究与伦理审查[J]. 中国医学伦理学, 2010, 23(4): 82-83.
- [2]World Medical Association. Declaration of Helsinki; recommendations guiding physicians in biomedical research in human subjects[Z]. Helsinki: World Medical Association, 1964 (Amended Tokyo 1975, Venice 1983, Hongkong 1989, Somerset West 1996, Edinburgh 2000).
- [3]李睿,陆芳,翁维良. 中医临床研究设计中的伦理问题与优化[J]. 中华中医药杂志, 2010, 25(7): 1050-1053.
- [4]王丽萍,王小丹. 药物临床试验伦理委员会运作的实践经验[J]. 药学服务与研究, 2012, 12(3): 191-220.
- [5]胡晋红. 论医学研究知情同意的伦理审查[J]. 中国医学伦理学, 2014, 27(2): 180-182.

- [6]陈敏,谢俊强. 1080 份知情同意书填写缺陷分析[J]. 现代医院, 2012, 12(6): 119-120.
- [7]胡晋红. 医学研究知情同意过程及信息告知的伦理审查要点[J]. 药学服务与研究, 2014, 14(3): 166-169.
- [8]汪秀琴. 临床研究的伦理审查——跟踪审查[J]. 中国医学伦理学, 2011, 24(7): 677-678.
- [9]张玮静,陆琴,吴炎. 论伦理委员会秘书在临床研究跟踪审查中的作用[J]. 中国医学伦理学, 2014, 27(3): 327-329.
- [10]叶茂林. 科技评价理论与方法[M]. 北京: 社会科学文献出版社, 2007: 188-200.
- [11]杜栋,庞庆华,吴炎. 现代综合评价方法与案例精选[M]. 2 版. 北京: 清华大学出版社, 2008: 6-8.
- [12]付赞,方德英. 雷达图法在综合评价中的应用研究[J]. 统计与决策, 2007(24): 176-178.
- [13]吴大嵘,梁伟雄,温泽淮,等. 建立中风病血瘀证宏观辩证量化标准的方法探讨[J]. 广州中医药大学学报, 1999, 4(16): 249-258.
- [14]邬新峰,赖世隆,梁伟雄. 中医药临床疗效评价中结局指标的选择与应用[J]. 广州中医药大学学报, 2002, 19(4): 251-255.
- [15]韩小强,张京萍,席家宁医学伦理审查在医院的规范运行研究[J]. 中国医药, 2012, 7(5): 642.

(2014-05-23 收稿 责任编辑: 徐颖)