

思路与方法

基于数据挖掘的高脂血症中医处方用药规律研究

吴嘉瑞 蔺梦娟 张晓朦 张 冰 盛晓光

(北京中医药大学, 北京, 100102)

摘要 目的:分析高脂血症的中医处方用药规律。方法:收集治疗高脂血症的中医处方,在建立处方数据库的基础上,采用关联规则 Apriori 算法和复杂系统熵聚类等方法开展研究,确定处方中各种药物的使用频次及药物之间的关联规则等。结果:高频次药物包括泽泻、山楂、丹参、决明子、茯苓等;高频次药物组合包括“山楂,泽泻”“山楂,丹参”“丹参,泽泻”等;置信度较高的关联规则包括“何首乌→山楂”“丹参,何首乌→山楂”“何首乌,泽泻→山楂”等;新处方包括“枳壳,延胡索,瓜蒌皮,青皮”“熟地黄,决明子,山药,何首乌,大黄”等。结论:处方用药多具健脾利湿化痰、活血化瘀清热之功,体现了中医痰瘀同治,标本兼顾的诊疗原则。

关键词 高脂血症;用药规律;关联规则;熵聚类

Analysis on Medication Rules of TCM Prescriptions for Hyperlipidemia Based on Data Mining

Wu Jiarui, Lin Mengjuan, Zhang Xiaomeng, Zhang Bing, Sheng Xiaoguang

(Beijing University of Chinese Medicine, Beijing 100102, China)

Abstract Objective: To analyze the medication rules of prescriptions for hyperlipidemia. **Methods:** Based on the database built by the prescriptions for hyperlipidemia, the methods of association rules with apriori algorithm and complex system entropy cluster were used to achieve the frequency of medicines, association rules between drugs and others. **Results:** The data-mining results indicated that the highest frequency used Chinese medicine were *Alismatis Rhizoma*, *Crataegi Fructus*, *Salviae Miltiorrhizae Radix et Rhizoma*, *Cassiae Semen*, and *Poria*. The most frequent drug combinations were “*Crataegi Fructus*, *Alismatis Rhizoma*”, “*Crataegi Fructus*, *Salviae Miltiorrhizae Radix et Rhizoma*”, and “*Salviae Miltiorrhizae Radix et Rhizoma*, *Alismatis Rhizoma*”. The drugs with a high degree confidence coefficient of association rules included “*Polygoni Multiflori Radix* - > *Crataegi Fructus*”, “*Salviae Miltiorrhizae Radix et Rhizoma*, *Polygoni Multiflori Radix* - > *Crataegi Fructus*”, and “*Polygoni Multiflori Radix*, *Alismatis Rhizoma* - > *Crataegi Fructus*”. The new prescriptions included “*Aurantii Fructus*, *Corydalis Rhizoma*, *Trichosanthis Pericarpium*, *Citri Reticulatae Pericarpium Viride*”, and “*Rehmanniae Radix Praeparata*, *Cassiae Semen*, *Dioscoreae Rhizoma*, *Polygoni Multiflori Radix*, *Rhei Radix et Rhizoma*”. **Conclusion:** The Chinese medicine applied in the prescriptions for hyperlipidemia mostly have the effects of invigorating the spleen, eliminating dampness, reducing phlegm, promoting blood circulation by removing blood stasis and clearing heat, which reflects the principles of synchronically treating phlegm and blood stasis, giving consideration to both the incidental and fundamental.

Key Words Hyperlipidemia; Medication rules; Association rules; Clustering algorithm

中图分类号:R289 文献标识码:A doi:10.3969/j.issn.1673-7202.2015.06.029

高脂血症是由于脂肪代谢或转运异常使血浆中一种或几种脂质高于正常范围的一种常见病,以血清中的总胆固醇(TC)、三酰甘油(TG)、低密度脂蛋白胆固醇(LDL-C)的升高,高密度脂蛋白胆固醇(HDL-C)的降低为特征^[1-2]。中医认为,高脂血症病因病机错综复杂,与饮食不节、先天禀赋不足、年老体衰或思虑过度、情志内伤等所致肝、脾、肾三脏

功能失调、代谢障碍、水湿痰饮瘀血病理产物的形成关系密切,为本虚标实、虚实夹杂之证,肝、脾、肾三脏之虚为本,痰浊、瘀血为标^[1,3-4]。本研究收集中医治疗高脂血症处方 98 首,采用关联规则、复杂系统熵聚类等方法,分析处方中药物的使用频次及药物之间的关联规则、处方规律,探讨中医治疗高脂血症的用药特点和学术思想。

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(1977.1—),男,博士,副教授,硕士研究生导师,从事临床中药学研究,Tel:(010)84738662,E-mail:exogamy@163.com

通信作者:张冰(1959.8—),女,教授,博士,博士研究生导师,从事临床中药学研究,Tel:(010)84738662,E-mail:zhangbing@263.net

1 资料与方法

1.1 处方来源与筛选 本研究以温武兵、刘鹏、李姝淳主编的《高脂血症百家百方》(中国中医药出版社,2012 年出版)^[1]为处方来源,共遴选摘录高脂血症中医处方 98 首。

1.2 分析软件 本研究采用“中医传承辅助系统(V2.0)”软件^[5],由中国中医科学院中药研究所提供。

1.3 处方的录入与核对 将上述筛选出的处方录入“中医传承辅助平台(V2.0)”中,录入完成后,由双人负责数据的审核,以确保数据的准确性。

1.4 数据分析

1.4.1 药物频次统计分析 将 98 首高脂血症中医处方中每味中药的出现频次从高到低排序,并将统计结果列出。

1.4.2 组方规律分析 在支持度 $\geq 10\%$,置信度 ≥ 0.9 条件下进行用药模式和关联规则分析,按照药物组合出现频次从高到低和药物关联规则置信度由高到低的顺序进行排序,列出高频次药物组合和药物组合关联规则结果;并分别在支持度为 10%、20%、30%,置信度 ≥ 0.9 条件下进行高频药物组合网络可视化展示^[6-8]。

1.4.3 新方分析 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 2,进行聚类分析(核心算法包括改进的互信息法和复杂系统熵聚类方法),分析获得药物间关联度,并按关联系数由高到低的顺序列出结果,同时运算得核心组合与新处方,实现新处方的网络可视化展示^[6-9]。

表 1 药物频次统计表

序号	中药名称	频次	序号	中药名称	频次
1	泽泻	66	17	法半夏	13
2	山楂	61	18	蒲黄	13
3	丹参	46	19	白芍	12
4	决明子	33	20	制何首乌	12
5	茯苓	30	21	枸杞子	12
6	白术	27	22	葛根	11
7	何首乌	26	23	荷叶	11
8	川芎	20	24	水蛭	11
9	黄芪	20	25	瓜蒌	11
10	当归	18	26	虎杖	11
11	陈皮	17	27	红花	11
12	甘草	15	28	党参	10
13	大黄	15	29	生地黄	10
14	茵陈	14	30	柴胡	10
15	郁金	14	31	桑寄生	10
16	半夏	14	32	赤芍	10

2 结果

2.1 用药频次统计 统计结果显示,98 首处方中共包含 190 味药物,将药物按使用频次从高到低进行排序,前三位分别是泽泻(66 次)、山楂(61 次)和丹参(46 次),出现频次 ≥ 10 次的中药见表 1。依据《中药学》教材^[10]对表 1 所示药物药性进行统计分析,结果显示:1)四气:寒(15.6%),微寒(21.9%),平(25.0%),微温(12.5%),温(21.9%);2)五味:苦(56.3%),甘(53.1%),辛(31.3%),酸(6.3%),咸(6.3%)、淡(3.1%);3)归经:归肝经(62.5%),归脾经(43.8%),归胃经(28.1%),归肺经(25.0%),归肾经(21.9%),归心经(15.6%),归胆经(15.6%),归心包经(12.5%),归大肠经(9.4%),归膀胱经(3.1%)。

表 2 处方中高频次药物组合(支持度 $\geq 10\%$,置信度 ≥ 0.9)

序号	药物组合	频次	序号	药物组合	频次
1	山楂,泽泻	45	34	陈皮,茯苓	12
2	山楂,丹参	34	35	山楂,白术,茯苓	12
3	丹参,泽泻	33	36	山楂,陈皮	11
4	山楂,决明子	27	37	大黄,山楂	11
5	泽泻,茯苓	27	38	丹参,当归	11
6	山楂,何首乌	26	39	丹参,茯苓	11
7	山楂,丹参,泽泻	26	40	丹参,白术	11
8	决明子,泽泻	25	41	甘草,泽泻	11
9	白术,泽泻	20	42	半夏,泽泻	11
10	山楂,决明子,泽泻	20	43	泽泻,茵陈	11
11	白术,茯苓	18	44	甘草,茯苓	11
12	山楂,茯苓	17	45	山楂,川芎,泽泻	11
13	山楂,白术	17	46	丹参,泽泻,茯苓	11
14	丹参,决明子	17	47	山楂,丹参,决明子,泽泻	11
15	何首乌,泽泻	17	48	山楂,白术,泽泻,茯苓	11
16	山楂,何首乌,泽泻	17	49	山楂,当归	10
17	决明子,何首乌	16	50	山楂,桑寄生	10
18	山楂,泽泻,茯苓	16	51	山楂,茵陈	10
19	山楂,决明子,何首乌	16	52	川芎,丹参	10
20	白术,泽泻,茯苓	16	53	丹参,郁金	10
21	山楂,黄芪	14	54	郁金,泽泻	10
22	川芎,泽泻	14	55	制何首乌,泽泻	10
23	山楂,丹参,决明子	14	56	大黄,泽泻	10
24	山楂,白术,泽泻	14	57	泽泻,枸杞子	10
25	山楂,川芎	13	58	甘草,白术	10
26	丹参,何首乌	13	59	山楂,丹参,黄芪	10
27	丹参,黄芪	13	60	山楂,白术,陈皮	10
28	白术,陈皮	13	61	决明子,何首乌,泽泻	10
29	山楂,丹参,何首乌	13	62	甘草,泽泻,茯苓	10
30	丹参,决明子,泽泻	13	63	白术,陈皮,泽泻	10
31	当归,泽泻	12	64	白术,陈皮,茯苓	10
32	泽泻,黄芪	12	65	山楂,决明子,何首乌,泽泻	10
33	陈皮,泽泻	12			

2.2 基于关联规则^[11]的组方规律分析 研究结果显示,常用药物组合 65 个,其中 2 味药的组合有 43

个,3 味药的组合 19 个,4 味药的组合 3 个,出现频次最高的 3 个药物组合依次是“山楂、泽泻”(45 次)、“山楂、丹参”(34 次)、“丹参、泽泻”(33 次),具体见表 2,关联规则如表 3,支持度分别为 10%、20%、30% 条件下的药物组合网络图见图 1。

表 3 处方中高置信度关联规则(支持度≥10%,置信度≥0.9)

序号	规则	置信度
1	何首乌→山楂	1
2	丹参,何首乌→山楂	1
3	何首乌,泽泻→山楂	1
4	决明子,何首乌→山楂	1
5	丹参,茯苓→泽泻	1
6	山楂,茯苓→泽泻	0.941 176
7	山楂,白术,茯苓→泽泻	0.916 667
8	山楂,陈皮→白术	0.909 091
9	甘草,茯苓→泽泻	0.909 091
10	甘草,泽泻→茯苓	0.909 091

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	枳壳,蜈蚣	0.029 352	15	熟地黄,墨旱莲	0.026 18
2	枳壳,浙贝母	0.029 352	16	熟地黄,酸枣仁	0.026 18
3	枳壳,蝉蜕	0.029 352	17	熟地黄,木瓜	0.026 18
4	枳壳,天花粉	0.029 352	18	熟地黄,碧桃干	0.026 18
5	黄芩,麻黄	0.029 352	19	熟地黄,辛夷	0.026 18
6	黄芩,浙贝母	0.029 352	20	熟地黄,龟甲胶	0.026 18
7	黄芩,肉桂	0.029 352	21	熟地黄,阿胶	0.026 18
8	黄芩,滑石	0.029 352	22	熟地黄,鹿角胶	0.026 18
9	黄芩,薄荷叶	0.029 352	23	熟地黄,浮小麦	0.026 18
10	黄芩,天花粉	0.029 352	24	熟地黄,玉米须	0.026 18
11	黄芩,炮姜	0.029 352	25	熟地黄,龙骨	0.026 18
12	黄芩,山楂	0.028 733	26	黄芩,决明子	0.0261 41
13	桃仁,冬瓜仁	0.027 643	27	山楂,黄精	0.0250 15
14	桃仁,瓜蒌仁	0.027 643			

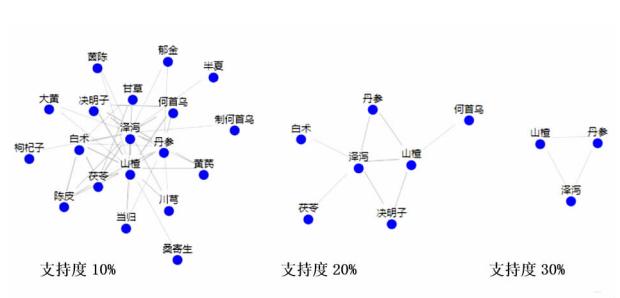


图 1 处方关联规则网络图

(支持度 10%、20%、30%,置信度≥0.9)

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
在相关度为 8,惩罚度为 2 的条件^[12]下,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 0.025 以上的药对列表,见表 4。

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(相关度为 8,惩罚度为 2),基于复杂系统熵聚类^[13-14],演化出 3~4 味药物核心组合 16 个,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 8 个新处方,具体见表 6,新处方网络图见图 2。

表 5 基于复杂系统熵聚类的药物核心组合

序号	核心组合	序号	核心组合
1	枳壳-延胡索-瓜蒌皮	9	枳壳-延胡索-青皮
2	熟地黄-决明子-山药	10	何首乌-决明子-大黄
3	黄芩-黄连-知母	11	黄芩-黄连-干姜
4	白芥子-莱菔子-焦神曲	12	白芥子-莱菔子-紫苏子
5	吴茱萸-牡丹皮-龙骨	13	熟地黄-吴茱萸-牡丹皮- 枸杞子-山药
6	桃仁-土鳖虫-浮小麦	14	桃仁-赤芍-红花
7	丹参-僵蚕-杜仲	15	丹参-桑寄生-杜仲
8	甘草-陈皮-白术-制何首乌	16	甘草-半夏-陈皮-白术-茯苓

表 6 基于熵层次聚类的新处方

序号	新方组合
1	枳壳,延胡索,瓜蒌皮,青皮
2	熟地黄,决明子,山药,何首乌,大黄
3	黄芩,黄连,知母,干姜
4	白芥子,莱菔子,焦神曲,紫苏子
5	吴茱萸,牡丹皮,龙骨,熟地黄,枸杞子,山药
6	桃仁,土鳖虫,浮小麦,赤芍,红花
7	丹参,僵蚕,杜仲,桑寄生
8	甘草,陈皮,白术,制何首乌,半夏,茯苓

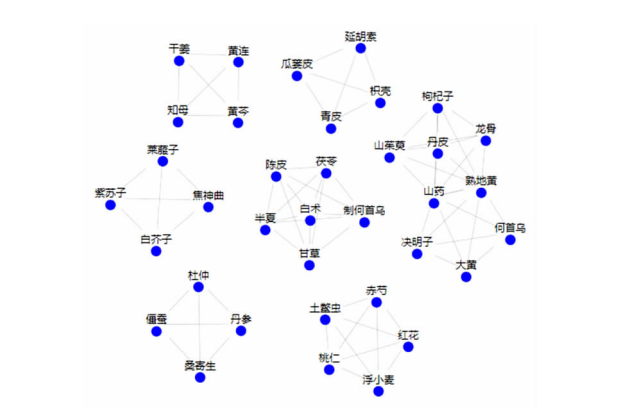


图 2 新方分析网络展示图

3 讨论

本研究运用关联规则和聚类算法分析了中医治疗高脂血症处方的用药规律。高频次药物包括泽泻(66 次)、山楂(61 次)和丹参(46 次)等。从功效角度分析,表 1 中所示药物多具有活血化瘀(34.4%)、利湿化痰(37.5%)、补血益气(37.5%)和疏肝理气(34.4%)等^[10];从药理角度分析,这些药物多具有降血脂(62.5%),抗凝血、预防血栓形成(40.6%),降血压(53.1%),扩张冠脉、增加冠脉血流量、提高心肌耐缺氧能力(40.6%),扩张脑动脉、增加脑血流

量(18.8%),保肝(43.8%),利胆(25.0%)等作用^[15]。这一结果表明,中医在临床实践中,既注重中医药理论指导,又借鉴了现代医药的研究成果;另一方面,亦可见中医药治疗高脂血症能够有效预防和治疗其多种兼症和继发病症,彰显中医药整体观念、辨证论治思想和中药多器官、多靶点作用特点的独特优势。

关联规则研究显示,置信度为1的关联规则有:“何首乌→山楂”“丹参,何首乌→山楂”“何首乌,泽泻→山楂”“决明子,何首乌→山楂”“丹参,茯苓→泽泻”等。如图1所示,在支持度较低(10%)时,可全面展示药物的使用状况,在支持度上升到20%时,可清晰展示出高脂血症中医处方中的关键药物及其关系,其中泽泻主归肾经,利尿渗湿行痰饮,泄热泄浊以升清阳;茯苓、白术,燥湿利水,健脾益气,扶正祛邪,为生痰之源(脾)而设,三者协同,使湿无所聚,痰无由生;丹参、山楂归肝经,活血祛瘀,通利血脉;山楂消食化积,决明子、何首乌润肠通便,三者消降结合以通腑泄浊;诸药合用,标本兼顾,痰瘀同治,泄浊升清,使浊脂浊膏无所由生,高脂血症渐愈^[10]。当支持度上升到30%时,可明确得出“泽泻-山楂-丹参”的组合,此三味药组合肝脾肾三脏共调,水湿痰饮瘀血同治,以治标为主。高脂血症因虚致实而产生的水湿痰饮瘀血等病理产物反过来也会成为致病因素,甚至引发其他继发病证,故患者无论处于高脂血症哪一阶段,治标是必须的,尤其处于高脂血症初、中期,正气尚足,以祛邪治标为主。

综上所述,本研究运用关联规则和聚类算法分析了中医治疗高脂血症的用药规律,希望本文对高脂血症的临床治疗、实验室研究、中药降血脂活性成分的发现及新药研发有一定的借鉴意义。当然,应用数据挖掘得到的组方规则、核心组合和新处方等还需进一步临床验证,数据挖掘方法也有其局限性,

需结合中医药理论与临床实践综合分析、评价。

参考文献

- [1]温武兵,刘鹏,李姝淳.高脂血症百家百方[M].北京:中国中医药出版社,2012.
- [2]杨永磊,谷新顺.血脂异常诊断与治疗进展[J].中国误诊学杂志,2011,11(3):525-526.
- [3]王喜福,祖麟麟,王成钢,等.复方丹参滴丸对高脂血症患者红细胞变形性的影响及其机制分析[J].中国医药,2013,8(4):456.
- [4]章赛月,朱晓岚.中医体质分型与高脂血症患者血脂水平的相关性研究[J].中国中医药科技,2013,20(3):217,232.
- [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]吴嘉瑞,郭位先,张晓滕,等.基于数据挖掘的孟河名医马培之治疗咳嗽用药规律研究[J].中国中医药信息杂志,2014,21(1):13-16.
- [7]吴嘉瑞,童有健,张晓滕,等.基于关联规则和复杂系统熵聚类的邓星伯治疗肺系病证用药规律研究[J].中国实验方剂学杂志,2014,20(7):223-226.
- [8]吴嘉瑞,张冰,杨冰,等.基于关联规则和复杂系统熵聚类的颜正华教授治疗呃逆用药规律研究[J].中华中医药杂志,2013,28(11):3416-3419.
- [9]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.
- [10]高学敏.中药学[M].2版.北京:中国中医药出版社,2007.
- [11]刘亚波.关联规则挖掘方法的研究及应用[D].长春:吉林大学,2005.
- [12]杨洪军,唐仕欢,卢朋.中医传承辅助平台的开发与应用[M].福州:福建科学技术出版社,2013:47-49.
- [13]杨洪军,赵亚丽,刘艳骄,等.“熵”在中医方证研究中的运用[J].中国中医基础医学杂志,2004,10(9):16-19.
- [14]吴嘉瑞,郭位先,张冰,等.基于关联规则和复杂系统熵聚类的颜正华治疗气滞证用药规律研究[J].中国中医基础医学杂志,2013,19(9):1081-1083.
- [15]彭成.中药药理学[M].3版.北京:中国中医药出版社,2012.

(2014-07-01 收稿 责任编辑:徐颖)