

• 循证医学与临床用药 •

基于数据挖掘的冠心病辨证用药规律研究

梁 杨^{1,2,3}, 丁长松^{1,2*}

1. 湖南中医药大学信息科学与工程学院, 湖南 长沙 410208

2. 湖南省中医药大数据分析实验室, 湖南 长沙 410208

3. 中南大学计算机学院, 湖南 长沙 410000

摘要: **目的** 基于数据挖掘技术研究分析冠心病辨证用药规律。**方法** 首先搜集整理名家医案处方, 对数据进行规范化处理, 构建冠心病处方数据库; 再借助统计学软件 IBM SPSS Statistics 20.0 对药物频次、病位证素频次及病性证素频次的分布进行统计; 最后应用 Apriori 关联规则算法等进行数据挖掘, 分析潜藏的用药配伍规律。**结果** 对遴选出的 145 首处方进行分析, 涉及 216 味常用中药、8 个病位证素和 12 个病性证素; 挖掘出用药频次大于平均频次的药物 46 味, 使用频次前 5 位的中药分别为丹参、甘草、栝楼、半夏、川芎, 病位证素以心、肾、脾、肝为主, 病性证素以血瘀、气虚、痰、阳虚、气滞、阴虚为主; 当满足最小支持度为 15% 且最小置信度为 70% 时, 关联规则分析结果显示共得到药对关联规则 15 条, 3 味药组关联规则 26 条, 其中置信度前 3 位的药对规则分别为薤白→栝楼、五味子→麦冬、郁金→丹参, 置信度前 3 位的 3 味药组关联规则分别为薤白+半夏→栝楼、丹参+薤白→栝楼、薤白+甘草→栝楼。**结论** 中医治疗冠心病辨证用药主以活血化瘀、行气导滞、补气益气、涤痰行气为法, 与其主要病性证素血瘀、气虚等相符, 药物大多归心经, 与其主要病位证素相一致。基于数据挖掘技术分析冠心病组方用药规律, 对于临床用药指导与应用具有重要价值。

关键词: 冠心病; 数据挖掘; Apriori 算法; 关联规则; 证素辨证; 丹参; 甘草; 栝楼; 半夏; 川芎

中图分类号: R28

文献标志码: A

文章编号: 0253-2670(2020)14-3742-05

DOI: 10.7501/j.issn.0253-2670.2020.14.018

Research on rules of syndrome differentiation and medication of coronary heart disease based on data mining

LIANG Yang^{1,2,3}, DING Chang-song^{1,2}

1. School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, China

2. TCM Big Data Analysis Laboratory of Hunan, Changsha 410208, China

3. School of Computer Science and Engineering, Central South University, Changsha 410000, China

Abstract: **Objective** To analyze the rules of coronary heart disease syndrome differentiation based on the data mining technology. **Methods** First of all, the famous medical records and prescriptions were collected and normalized to construct the database of coronary heart disease prescriptions. Then, IBM SPSS Statistics 20.0, a statistical software, was used to conduct statistics on the distribution of drug frequency, disease site syndrome frequency and disease syndrome frequency. Finally, the Apriori algorithm was applied to carry out data mining and analyze the underlying law of drug compatibility. **Results** A total of 145 prescriptions were selected and analyzed, including 216 Chinese medicines, eight syndromes of disease location and 12 syndromes of disease nature. Forty-six herbs with higher frequency were found, and the top five TCM herbs were *Salvia miltiorrhiza*, *Glycyrrhiza uralensis*, *Trichosanthes kirilowii*, *Pinellia ternata* and *Ligusticum chuanxiong*. The main syndromes of disease location were heart, kidney, spleen and liver, while the main syndromes of disease nature were blood stasis, *qi* deficiency, phlegm, deficiency of *yang*, *qi* stagnation, and deficiency of *yin*. When the minimum support was 15% and the minimum confidence coefficient was 70%, the association rule analysis results showed that a total of 15 association rules for two-herb-combinations and 26 association rules for the trigeminy

收稿日期: 2019-11-08

基金项目: 国家重点研发计划 (2017YFC1703306); 湖南省教育厅科学研究项目 (19C1391); 湖南省重点研发计划 (2017SK2111); 湖南省教育厅重点项目 (18A227); 湖南省自然科学基金 (2018JJ2301); 湖南省中医药科研计划重点课题 (2020002); 湖南中医药大学电子科学与技术学科开放基金 (2018DK04)

作者简介: 梁 杨, 男, 讲师, 博士生, 主要从事中医药信息化研究。E-mail: liangyang1987@tom.com

***通信作者:** 丁长松, 男, 教授, 博士生导师, 主要从事中医药信息化研究。E-mail: dinghongzhe@yeah.net

medications were obtained. The most frequency two-herb-combinations were *Allium macrostemon*→*T. kirilowii*, “*Schisandra chinensis*→*Ophiopogon japonicus*”, “*Curcumae Radix*→*S. miltiorrhiza*”. The most frequently used trigeminy medications were *A. macrostemon* + *P. ternata*→*T. kirilowii*, *S. miltiorrhiza* + *A. macrostemon*→*T. kirilowii* and *A. macrostemon* + *G. uralensis*→*T. kirilowii*. **Conclusion** The syndrome differentiation and medication of TCM in the treatment of coronary heart disease mainly focus on invigorating the circulation of blood and resolving stasis, moving *qi*, nourishing and replenishing *qi*, clearing up phlegms. It is consistent with the main syndromes of the disease nature, such as blood stasis and *qi* deficiency, and the herbs mostly return to the heart meridian, which is consistent with the main syndromes of the disease location. The rules of syndrome differentiation and medication of based on the data mining technique have great value for clinical medication guidance and application to analyze the prescription for coronary heart disease.

Key words: coronary heart disease; data mining; Apriori algorithm; association rules; syndrome differentiation; *Salvia miltiorrhiza* Bunge; *Glycyrrhiza uralensis* Fisch.; *Trichosanthes kirilowii* Maxim.; *Pinellia ternate* (Thunb.) Breit.; *Ligusticum chuanxiong* Hort.

冠心病是指血液中代谢不正常的脂质沉着在光滑的动脉内膜上,形成粥样化的白色斑块,也被称为冠状动脉粥样硬化性心脏病,其发病率在心血管疾病中一直处于前列,具有发病率高、致死率高、致残率高、复发率高及并发症多的“四高一多”特点^[1-2]。冠心病对人体健康危害巨大,据调查,截止到 2016 年冠心病死亡人数高达 173.6 万例,已成为心血管疾病中死亡人数仅次于脑卒中的第 2 大疾病^[3]。因此,为有效应对冠心病防治带来的严峻挑战,本研究通过对 145 例冠心病医案辨证数据整理,采用 IBM SPSS Statistics 20.0 统计学软件对药物频率分布进行统计,并应用 Apriori 关联规则算法进行挖掘分析,讨论冠心病用药配伍规律^[4-5],以期临床辨证中减少人为失误提供可靠支撑,为中医药治疗冠心病组方用药提供参考,从而更好地指导临床实践。

1 资料

1.1 资料来源

本研究中的冠心病中医医案处方以《冠心病百家百方》^[6]、《内科医案(上册)》^[7]、《大国医经典医案诠释:冠心病(病症篇)》^[8]为主要来源,共遴选摘录 145 首。

1.2 纳入标准

①所有诊断为冠心病患者,诊断符合国际心脏病学会和协会/世界卫生组织临床命名标准化专题组的联合报告《缺血性心脏病的命名及诊断标准》、中华人民共和国国家标准《中医临床诊疗术语》(证候部分)等有关公认的诊断标准,年龄、性别不限;②使用中药内服法治疗者,患者证型及用药记录等信息记录完整者;③同一患者复诊所开新药方视为可用信息。

1.3 排除标准

①中医临床诊断非冠心病的医案;②相关记录

不完整者,使用针灸、推拿、穴位贴敷等治法治疗者。

2 方法

2.1 数据规范化

2.1.1 中药名称规范化 根据《中药大辞典(第二版)》^[9]对中药名称进行规范,考虑到因医家地域或者习惯的差别,同一种药经常会出现名称不一致问题,为方便数据处理,本研究对中药名称规范化说明如下:(1)中药名称主要从别名、错别字、省略词等方面做出统一规范;(2)暂不考虑炮制方法和药物产地对中药药性的影响。部分中药名称规范化结果见表 1。

表 1 部分中药名称规范化

Table 1 Partial Chinese medicine names normalization

规范前	规范后	规范前	规范后
苡仁	薏苡仁	象贝母、川象贝	象贝
香谷芽	谷芽	紫丹参	丹参
青龙齿	龙齿	南沙参	沙参
附子	附片	川军	大黄
川陈皮、广陈皮	陈皮	潞党参	党参
瓜蒌	栝楼	萆薢	萆拔
熟地黄	熟地	元胡、玄胡	延胡索
川朴	厚朴	茯苓神	茯苓+茯神
川桂枝	桂枝	龙牡	龙骨+牡蛎
大生地、生地黄	生地	乳没	乳香+没药
九节菖蒲	节菖蒲	二活	羌活+独活
桂圆肉	龙眼肉	川断	续断

2.1.2 证型规范化 根据医案辨证结果将证名进行证型规范化,其中“证型=病位证素+病性证素”,按照《证素辨证学》的规范证名进行整理和规范,具体依据如下:(1)证名中可以包容某些尚未列入的证素;(2)可以在规范证名中加“兼/夹 XX”; (3)

根据中医学理论,有机组成证名。部分中医证型规范化后结果见表 2。

2.2 数据录入及核对

将遴选规范化的医案处方通过 Excel 2010 构建冠心病方剂数据库,采用双人分别录入模式,录入完成后进行双人交叉核对,若存在不一致处则协商解决,以确保数据的准确性和规范性。

表 2 部分中医证型规范化

Table 2 Partial TCM syndromes normalization

证名	病位证素	病性证素
肝火炽盛证	肝	热
心气虚证	心	气虚
心脾两虚证	心+脾	气虚+血虚
心阳气虚证	心	阳虚+气虚
心肾阴虚阳亢证	心+肾	阴虚+阳亢
心气阴虚证	心	气虚+阴虚+血虚
痰阻心脉证	心	痰
心脾气血虚证	心+脾	气虚+血虚
痰湿瘀阻心脉证	心	痰+血瘀
心阴阳气血虚证	心	气虚+血虚+阳虚+阴虚
心肾阴虚阳亢证	心+肾	阴虚+阳亢+(血虚)
心阳虚,痰湿内蕴证	心	痰+血瘀+气滞+阳虚
心阳气虚证	心	阳虚+气虚+寒
心脉瘀滞证	心	气滞+血瘀
心肾阳虚证	心+肾	阳虚
胆胃不和证	胆+胃	气滞+气虚
脾胃气虚证	脾+胃	气虚
肝郁气滞证	肝	气滞
心肾气阴虚证	心+肾	气虚+阴虚
肝血瘀证	肝	血瘀
心肾阳虚水泛证	心+肾	水停+阳虚

2.3 数据分析

描述性分析。采用 IBM SPSS Statistics 20.0 分别对药物和证素的频次与频率进行统计分析,并根据频率分布进行降序排序,计算每味药出现的平均频次,重点研究出现频次大于平均频次的药物。

关联规则分析。应用 Apriori 关联规则算法挖掘与分析药对配伍规律,从而寻找中医药治疗冠心病的用药模式。设置支持度 $\geq 15\%$ 、置信度 $\geq 70\%$,用于分别提取药物 2 项和 3 项关联规则结果,从而得到中医药治疗冠心病的常见药对和药组。

3 结果

3.1 药物频次、频率分析

本研究录入的 145 个冠心病医案处方中,其中涉及 216 味中药,累计频次 1 931 次,平均频次为 9 次,使用频次小于等于 9 次的中药共 170 味(506 次),占比为 26.2%,频次大于 9 次的中药 46 味(1 425 次),占比为 73.8%。出现频次大于平均频次的冠心病药物频次分布见表 3,冠心病药物使用频次整体分布见图 1。

取表 3 中用药频次在前 10%作为高频药物,分别为丹参、甘草、栝楼、半夏、川芎。

3.2 证素频次、频率分析

本研究录入的 145 个冠心病医案处方中,共涉及 8 个病位证素(心、肾、脾、肝、胃、胆、肺、心神)和 12 个病性证素(血瘀、气虚、痰、阳虚、

表 3 大于平均频次的冠心病药物频次分布

Table 3 Above-average usage frequency distribution table of coronary heart disease herbs

中药名称	频次	频率/%	中药名称	频次	频率/%
丹参	86	59.3	延胡索	27	18.1
甘草	79	54.5	太子参	27	18.1
栝楼	68	46.9	砂仁	23	15.7
半夏	65	44.8	枳实	21	14.7
川芎	56	38.6	香附	20	13.8
薤白	56	38.6	柴胡	19	13.1
当归	54	37.2	竹茹	19	13.1
茯苓	50	34.5	降香	18	12.4
红花	48	33.1	远志	17	11.7
黄芪	45	31.0	桔梗	17	11.7
赤芍	44	30.3	牛膝	16	11.0
麦冬	42	29.0	柏子仁	15	10.3
枳壳	41	28.3	地黄	15	10.3
党参	38	26.2	附片	14	9.7
桂枝	34	23.4	姜	14	9.7
五味子	33	22.8	人参	13	9.0
白芍	32	22.1	山楂	12	8.3
郁金	32	22.1	何首乌	12	8.3
桃仁	32	22.1	西洋参	12	8.3
酸枣仁	30	20.7	葛根	11	7.6
陈皮	30	20.7	夜交藤	11	7.6
白术	29	20.0	檀香	11	7.6
三七	27	18.1	橘红	10	6.9

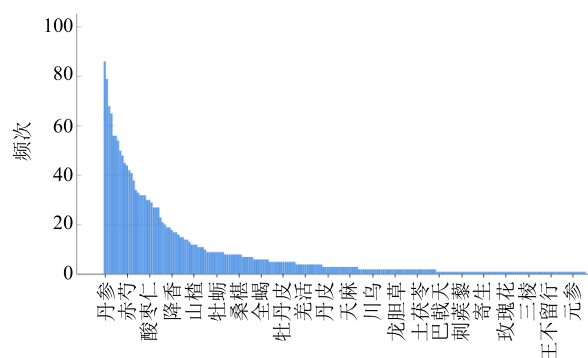


图 1 冠心病药物使用频次整体分布

Fig. 1 Overall distribution diagram of coronary heart disease herb usage frequency

气滞、阴虚、血虚、湿、寒、水停、阳亢、热)。

3.2.1 病位证素 冠心病病位证素累计频次 179 次, 则冠心病病位证素频次分布见表 4, 冠心病病位证素频次分布见图 2。

表 4 冠心病病位证素频次分布

Table 4 Frequency distribution table of disease location of coronary heart disease

病位证素	频次	频率/%	病位证素	频次	频率/%
心	142	97.9	胃	3	2.3
肾	16	11.0	胆	2	1.6
脾	9	6.2	肺	1	0.7
肝	5	3.4	心神	1	0.7

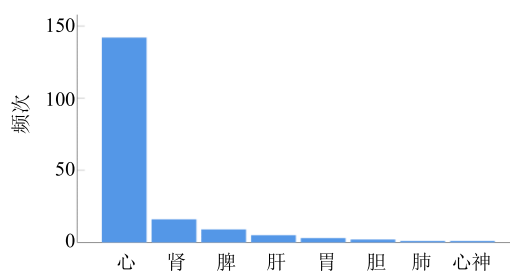


图 2 冠心病病位证素频次分布

Fig. 2 Frequency distribution diagram of disease location of coronary heart disease

冠心病病位证素平均频次为 22 次, 频次较高的病位证素依次为心、肾、脾、肝。

3.2.2 病性证素 冠心病病性证素累计频次 359 次, 则冠心病病性证素频次分布见表 5, 冠心病病性证素频次分布见图 3。

冠心病病性证素平均频次为 30 次, 因此, 频次较高的病性证素为血瘀、气虚、痰、阳虚、气滞、阴虚。

表 5 冠心病病性证素频次分布

Table 5 Frequency distribution table of disease nature of coronary heart disease

病性证素	频次	频率/%	病性证素	频次	频率/%
血瘀	84	58.0	血虚	16	11.0
气虚	79	54.5	湿	12	8.3
痰	52	35.9	寒	4	2.8
阳虚	37	25.5	水停	4	2.8
气滞	33	22.8	阳亢	3	2.1
阴虚	33	22.8	热	2	1.4

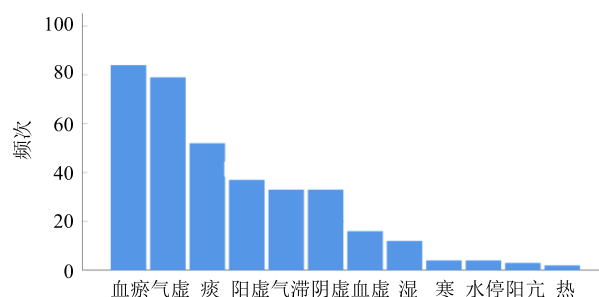


图 3 冠心病病性证素频次分布

Fig. 3 Frequency distribution diagram of disease nature of coronary heart disease

3.3 方药关联规则结果与分析

所谓关联规则, 主要含义为前项药物出现时, 后项可能会出现的药物, 可表示为“前项→后项”形式。支持度表示前项药物和后项药物在所有方剂中同时出现的概率; 置信度含义为在前项药物出现的处方中, 后项药物出现的概率, 用于衡量关联规则的可信程度。基于以上原则, 可得出中医药治疗冠心病常见的药对、药组及中药之间的潜在共性规律。冠心病药对关联规则和冠心病 3 味药组关联规则分别见表 6、7。

4 讨论

冠心病中出现频率较高的药物为: 丹参、甘草、栝楼、半夏、川芎; 病性证素主要为: 血瘀、气虚、痰、阳虚、气滞、阴虚; 病位证素主要为: 心、肾、脾、肝。

4.1 含丹参的方药用药规律研究

丹参是治疗冠心病的方药中最为常见的活血化瘀药, 且具有止痛效果, 又因其归心、肝经, 作为出现频率最高的药物, 其常见的药物组合有“黄芪→丹参”“郁金→丹参”“酸枣仁→丹参”“党参→丹参”“川芎+栝楼→丹参”“栝楼+川芎→丹参”。

表 6 冠心病药对关联规则

Table 6 Association rules table of two-tuples for herb of coronary heart disease

关联规则	置信度/%	关联规则	置信度/%
桃仁→红花	71.9	赤芍→川芎	70.5
白术→茯苓	75.9	黄芪→丹参	77.8
酸枣仁→丹参	73.3	郁金→丹参	87.5
党参→丹参	73.7	薤白→栝楼	96.4
栝楼→薤白	79.4	陈皮→半夏	80.0
茯苓→半夏	72.0	党参→甘草	73.7
五味子→甘草	81.8	五味子→麦冬	90.9
麦冬→五味子	71.4		

表 7 冠心病 3 味药组关联规则

Table 7 Association rules table of three-tuples for herb of coronary heart disease

关联规则	置信度/%	关联规则	置信度/%
当归+川芎→红花	73.3	红花+丹参→川芎	71.0
红花+赤芍→川芎	88.5	赤芍+川芎→红花	74.2
川芎+栝楼→丹参	76.7	薤白+川芎→栝楼	96.0
栝楼+川芎→薤白	80.0	栝楼+川芎→丹参	76.7
薤白+川芎→栝楼	96.0	川芎+栝楼→薤白	80.0
丹参+薤白→栝楼	97.1	丹参+栝楼→薤白	75.0
丹参+茯苓→半夏	76.7	栝楼+甘草→半夏	74.3
薤白+半夏→栝楼	100.0	栝楼+半夏→薤白	82.2
薤白+甘草→栝楼	96.9	栝楼+甘草→薤白	88.6
栝楼+甘草→半夏	74.3	茯苓+薤白→栝楼	91.7
茯苓+栝楼→薤白	84.6	薤白+甘草→半夏	75.0
茯苓+甘草→半夏	71.9	麦冬+甘草→五味子	82.8
五味子+麦冬→甘草	80.0	五味子+甘草→麦冬	88.9

4.2 含甘草的方药用药规律研究

甘草作为补益类药，可用来补脾益气，是中药方剂中常用药物，因其归心、脾、胃经，可治疗心悸气短。常见的药物组合有“党参→甘草”“五味子→甘草”“五味子+麦冬→甘草”。

4.3 含栝楼的方药用药规律研究

栝楼以其化痰、宽胸理气的功效常出现于方药中。常见的药物组合有“薤白→栝楼”“薤白+川芎→栝楼”“薤白+川芎→栝楼”“丹参+薤白→栝楼”“薤白+半夏→栝楼”“薤白+甘草→栝楼”“茯苓+薤白→栝楼”。

4.4 含半夏的方药用药规律研究

半夏同栝楼一样均作为化痰药用于冠心病治疗中，可用于湿痰寒痰等情况，归脾、胃经，常见的

药物组合有“陈皮→半夏”“茯苓→半夏”“丹参+茯苓→半夏”“栝楼+甘草→半夏”“茯苓+甘草→半夏”“薤白+甘草→半夏”“栝楼+甘草→半夏”。

4.5 含川芎的方药用药规律研究

川芎是较为常见的活血止痛药，可用来活血行气，归肝、胆、心经。常见的药物组合有“赤芍→川芎”“红花+丹参→川芎”“红花+赤芍→川芎”。

冠心病主要病位证素为心，主要病性证素为实证血瘀和虚证气虚。其中，置信度前 3 位的药对规则分别为：“薤白→栝楼”“五味子→麦冬”“郁金→丹参”，置信度前 3 位的 3 味药组规则分别为：“薤白+半夏→栝楼”“丹参+薤白→栝楼”“薤白+甘草→栝楼”。其中，丹参可用于活血化瘀，甘草可用于补气，栝楼、半夏可用于化痰，川芎可用来活血行气。因此，中医治疗冠心病辨证用药主以活血化瘀、行气导滞、补气益气、涤痰行气为法，与其主要病性证素血瘀、气虚、痰、阳虚、气滞、阴虚等相符，药物大多归心经，与其主要病位证素相一致。

综上所述，本研究基于数据挖掘方法对冠心病医案处方的用药规律从药对、3 味药组等多个角度进行关联规则分析，挖掘了中医药治疗冠心病的组方配伍规律，并具体分析得出了针对频率较高药物、病性证素、病位证素的用药规律。本研究为深入挖掘冠心病等常见疾病的辨证用药规律，基于数据挖掘技术分析冠心病组方用药规律，方法科学有效，为临床辨证用药提供了可靠的依据，对于临床用药指导与应用具有重要价值，但所得结论仍需临床的检验。

参考文献

- [1] 于林童, 曲文白, 余新波, 等. 数据挖掘方法在名老中医用药规律研究中的应用现状 [J]. 中医杂志, 2017, 58(10): 886-900.
- [2] 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.
- [3] 王 朔, 解 静, 李 琳, 等. 防治冠心病中药质量标志物的研究进展 [J]. 中草药, 2019, 50(19): 4620-4627.
- [4] 梁 杨, 丁长松, 黄辛迪, 等. 大数据环境下的中医药数据管理策略 [J]. 中国中药杂志, 2018, 43(4): 840-846.
- [5] 梁 杨, 丁长松, 于俊洋. 基于 Hadoop 的中医药数据管理策略研究 [J]. 中国中医药信息杂志, 2018, 25(5): 96-100.
- [6] 汤晓龙. 冠心病百家百方 [M]. 北京: 中国中医药出版社, 2012.
- [7] 罗和古, 杜少辉, 曾令真, 等. 内科医案 (上册) [M]. 北京: 中国医药科技出版社, 2014.
- [8] 邓小英. 冠心病 (大国医经典医案诠释 (病症篇)) [M]. 北京: 中国医药科技出版社, 2016.
- [9] 南京中医药大学. 中药大辞典 [M]. 第 2 版. 上海: 上海科学技术出版社, 2006.