

· 学术探讨 ·

冠心病血瘀证症状与药物组成最大频繁关联模式挖掘

唐梅森^{1△} 龚后武¹ 黄政德¹ 晏峻峰¹ 李 杰²

摘要 收集中医治疗冠心病血瘀证医案 121 例,建立冠心病血瘀证诊疗数据库,运用数据挖掘探讨中医治疗冠心病血瘀证的症状与用药的内在规律,探索冠心病血瘀证症状与药物组成的最大频繁关联模式。在所整理的 121 例医案中,胸痛、胸闷及头痛是冠心病血瘀证患者的常见症状,活血药、化瘀药及行气药之间的配伍是最常用的药物组合,二者之间的关联构成了冠心病血瘀证常见的最大频繁关联模式。以关联模式挖掘所得的冠心病血瘀证症状与药物关联模式反应了治疗冠心病血瘀证标本同治的治疗思想。

关键词 冠心病;血瘀证;症状;中药;最大频繁关联模式

Maximal Frequent Association Patterns between Symptoms of Coronary Heart Disease Patients with Blood Stasis Syndrome and Medication Constitutions TANG Mei-sen¹, GONG Hou-wu¹, HUANG Zheng-de¹, YAN Jun-feng¹, and LI jie² 1 Graduate School, Hunan University of Chinese Medicine, Changsha (410208); 2 College of Traditional Chinese Medicine, Hunan University of Chinese Medicine, Changsha (410208)

ABSTRACT A total of 121 medical cases concerning treating coronary heart disease (CHD) patients with blood stasis syndrome (BSS) by Chinese Medicine were collected to establish a database for CHD patients with BSS. By using data mining authors tried to explore inherent laws of its symptoms and medication of Chinese Medicine, and to probe maximal frequent association patterns between its symptoms and medication constitutions. Of the 121 medical cases, chest pain, chest stiffness, and headache were their common symptoms. Compatibilities of blood-activating drugs, stasis-resolving drugs, and qi-promoting drugs were most commonly used. The association between symptoms and compatibilities constituted a most often seen maximal frequent association pattern, which reflected an idea of treating both principal and subordinate symptoms.

KEYWORDS coronary heart disease; blood stasis syndrome; symptom; Chinese materia medica; maximal frequent association pattern

冠心病(coronary heart disease, CHD)即冠状动脉粥样硬化性心脏病。中医学无冠心病这个病名。历代医籍中均有记载“卒心痛”、“厥心痛”、“胸痹心痛”、“真心痛”等病证的症状,与冠心病的患者表现的疾病特征是一致的。现代中医学用“胸痹”对冠心病辨

证施治。凡离开经脉的血液不能及时排出和消散,而停留于体内,或血液运行不畅,瘀积于经脉或脏腑组织器官之内的均称为瘀血。由瘀血内阻而引起的病证,称为血瘀证。以刺痛不移、拒按、肿块、出血、唇舌爪甲紫暗、脉涩等为辨证要点。主要临床表现:疼痛如针刺刀割,痛有定处而拒按,常在夜间加剧。以痛、紫、瘀、块、涩为特点。引起血瘀的原因有寒凝、气滞、气虚、外伤等。冠心病血瘀证是血瘀证研究中最活跃领域。“本虚标实,标实中血瘀贯穿发病过程始终”^[1],因此“活血化瘀”成为治疗冠心病血瘀证的基本思想,对患者具体辨证分型后,衍化出温阳活血、理气活血、益气活血、化浊活血等疗法,临床疗效显著。本研究运用数据挖掘探讨中医治疗冠心病血瘀证的症状与用药内在规律。

基金项目:国家自然科学基金资助项目(No.81001518);霍英东教育基金会高校青年教师基金(No.131042);中国博士后科学基金(No.2012M521526, No.2013T60771);湖南省科技厅计划项目(No.2013RS4038)

作者单位:1.湖南中医药大学研究生院(长沙 410208);2.湖南中医药大学中医学学院(长沙 410208)

通讯作者:李 杰, Tel:13308472992, E-mail:317768870@qq.com

△作者现在河南省平煤神马医疗集团总医院中西医结合科(河南 467000)

DOI: 10.7661/j. cjm. 20170206. 021

1 一般资料 121 例均为 2011 年 6 月—2012

年 6 月湖南中医药大学第一附属医院心内科门诊及住院 CHD 患者。CHD 诊断标准参照《缺血性心脏病的命名及诊断标准》^[2]。血瘀证和非血瘀证辨证参照《中医病症诊断疗效标准》^[3] 相关标准。纳入标准: 18~75 岁; 符合 CHD 诊断及辨证标准; 经伦理委员会同意, 签署医患沟通记录和知情同意书。排除标准: 甲状腺机能亢进性心肌病, 糖尿病性心肌病, 肺源性心肌病, 高血压性心肌病, 风湿性心脏病, 贫血性心脏病; 合并恶性肿瘤, 甲状腺疾病, 肝肾疾病, 结核等传染病; 资料不全。

按血瘀证常见的病因进行分型, 冠心病血瘀证临床依次常见以下兼证: (1) 气滞血瘀; (2) 阳虚血瘀 (3) 气虚血瘀; (4) 痰阻血瘀; (5) 寒凝血瘀; (6) 热结血瘀^[4]。

2 最大频繁关联模式挖掘

2.1 相关概念 列出相关最大频繁关联模式挖掘概念^[5]: 关联规则, 频繁项目集, 最大频繁集。

定义 1: 设 $I = \{I_1, I_2, \dots, I_m\}$ 是一个项目集合, 事务数据库 $D = \{D_1, D_2, \dots, D_n\}$ 是由一系列具有唯一标识的 TID 的事务组成, 每个事务 $t = \{I_{t1}, I_{t2}, \dots, I_{tk}\}$ 并且 $I_{ik} \in I$, 关联规则是形如 $X \Rightarrow Y$ 的蕴涵式, 其中 $X, Y \subset I$, 是两个项目集合, 成为项目集合并且 $X \cap Y = \emptyset$ 。

定义 2: 给定事务数据 D 和最小支持度书國值 minsupport, 对于项目集 $X \subseteq I$, 若 $\text{support}(X) \geq \text{minsupport}$, 则称 X 为 D 中的频繁项目集 (frequent itemsets, FI)。

定义 3: 给定事务数据 D 和最小支持度书國值 minsupport, 对于项目集 $X \subseteq I$, 若 $\text{support}(X) \geq \text{minsupport}$, 且对于 $\forall Y (Y \subseteq I \wedge X \subset Y)$, 都有 $\text{support}(Y) \geq \text{minsupport}$ 。则称 X 为 D 中的最大频繁项目集 (maximal frequent itemsets, MFI)。将所有最大频繁项目集组成的集合称为最大频繁集 (maximal frequent sets, MFS)。

2.2 算法的设计思想 算法首先是产生频繁 1-项集, 接着利用上述定理逐步的扩展频繁项集, 直至不能扩展 (即找到相应的最大项集) 为止^[6]。算法的运算步骤如下。

首先, 通过扫描数据库, 累计每个项的计数, 并收集满足最小支持度的项, 找出频繁 1-项集的集合, 该集合记作 L_1 。然后, L_1 用于找频繁 2-项集的集合 L_2 , L_2 用于找 L_3 。如此下去, 直到找不出 k -频繁项集为止。即通过逐次搜索各层的迭代方法, 即将 k -项集用于扩展 $(k+1)$ -项集, 来扩展数据集中的所有 FI。MFI 的所有子集即为 MFS。

2.3 算法伪代码描述

C_k : Candidate itemset of size k , L_k : frequent itemset of size k , $L_1 = \{\text{frequent items}\}$; for ($k = 1$; $L_k \neq \emptyset$; $k++$) do begin, C_{k+1} = candidates generated from L_k ; for each transaction t in database do, increment the count of all candidates in C_{k+1} that are contained in t , L_{k+1} = candidates in C_{k+1} with min_support, end, return $\bigcup_k L_k$ 。

3 冠心病血瘀证症状与药物组成最大频繁关联模式挖掘 (图 1, 表 1) 冠心病血瘀证症状与药物组成最频繁关联模式挖掘的流程: 该挖掘流程首先从临床医案数据中获取数据, 从中获取出现的冠心病血瘀证相关方剂。在获取到数据以后, 对数据进行预处理操作。对不规则的数据进行清理, 根据挖掘算法的特殊需要及出于对数据隐私保护的需要, 把有显示语义的字符串数据映射到语义隐含的数字串中, 以达到隐藏数据内容的目的。在数据预处理结束以后, 经过预处理后的数据被送到一个或多个具体数据挖掘算法进行挖掘操作。最后的挖掘结果可能被存储到某个存储系统中。

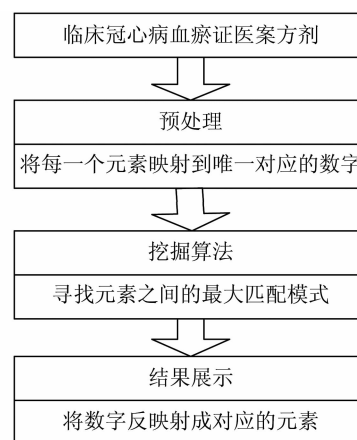


图 1 冠心病血瘀证症状与药物组成最大频繁关联模式挖掘流程

从 Excel 文件读入冠心病血瘀证方剂数据, 把读入的方剂药物组成部分进行基于正则表达式的预处理。原始数据的药物组成字段结果为 (药物名; 剂量; 单位), 由于研究药物名和症状的关系, 因此暂时把剂量和单位信息先去除, 把含有空字段的行过滤掉。对于输入规则要预先转变成数字 ID, 这样能够提高算法效率^[7], 先把症状名称及药物组成数字编码。症状名称和药物编码组成分别 1000、2000 开始编号。

4 冠心病血瘀证高频最大模式挖掘部分结果 (表 2) 执行挖掘算法, 寻找元素之间的最大匹配模

表 1 冠心病血瘀证原始数据示例

症状名称	药物组成
心胸阵痛如刺如绞,心悸,面色晦暗,舌暗,舌有瘀斑,舌下脉络青紫,脉细,脉涩,脉结,脉代	当归 9 g,生地黄 9 g,桃仁 12 g,红花 9 g,枳壳 6 g,赤芍 6 g,川芎 5 g,柴胡 3 g,桔梗 5 g,牛膝 9 g,甘草 3 g
心胸胀痛,常因情绪波动而加重,心悸,善太息,舌淡红,脉弦,脉细	柴胡 6 g,陈皮 6 g,川芎 4.5 g,香附 4.5 g,枳壳 4.5 g,芍药 4.5 g,炙甘草 1.5 g,五灵脂 15 g,生蒲黄 15 g,桃仁 15 g,红花 15 g,郁金 15 g,赤芍 15 g
心痛遇寒加剧得温痛减,形寒肢冷,面色苍白,舌淡暗,舌苔白,脉沉,脉迟,脉沉,脉紧	枳实 12 g,厚朴 12 g,薤白 9 g,桂枝 6 g,栝蒌 12 g

表 2 冠心病血瘀证高频最大模式挖掘部分结果

症状与药物组成规则	出现次数	相对支持度
胸痛(100) 头痛(201) 心悸(141) 桃仁(601) 红花(602)	121	0.15
胸刺痛(101) 脉涩(305) 桃仁(601) 红花(602) 赤芍(605)	118	0.15
胸痛(100) 舌质暗红(411) 舌有瘀斑(423) 川芎(604)	96	0.11
心悸(141) 胸闷(113) 畏冷肢凉(512) 黄芪(611) 桃仁(601) 红花(602)	103	0.13
胸闷(141) 心悸(113) 枯萎(610) 薤白(702) 茯苓(715)	72	0.11
胸闷(113) 咳痰(562) 枯萎(610) 桂枝(658) 茯苓(718)	69	0.11
胸胁胀痛(102) 柴胡(634) 羌活(684) 当归(603) 川芎(604)	84	0.11
心痛遇寒加重(142) 舌苔白腻(413) 脉沉(306) 丹参(742) 川芎(604)	58	0.10
心胸灼痛(143) 丹皮(745) 赤芍(605) 蒲黄(723)	61	0.10
胸闷(100) 胸痛(113) 畏冷肢凉(512) 附子(672) 川乌(625)	86	0.12

式。将结果所得 ID 映射成病案中对应的症状名称及药物名,进一步分析其临床指导意义。通过治疗冠心病血瘀证的临床医案献数据可以看出:(1)冠心病血瘀证中较为常见的证型为阳虚血瘀证和气滞血瘀证。(2)桃仁、红花、当归、川芎是常用的四味药,胸痛、头痛、胸闷是常见的三个症状,这可以分别从以上结果数据中得到印证。(3)川芎在临床中常用于气滞血瘀之胸胁、腹部诸疼痛症的治疗,起活血化瘀、行气止痛之效,且多以他药佐之。(4)桃仁和红花组成的药对和胸痛/胸闷这个症状组关系最密切。桃仁破血行滞而润燥,红花活血祛瘀以止痛,针对瘀血内阻胸部,气机郁结所致诸症。当归、川芎、赤芍都是活血的中药,并不是直接可以治疗胸痛、胸闷,而是由于能助君药活血祛瘀,所以在结果中当归、川芎、赤芍都是频繁出现的。

5 讨 论 关联规则分析就是从大量的数据中发现项集之间的关联、相关关系、因果结构以及项集的频繁模式。关联规则用于研究症状与药物药物内在规律,可以在不同层次发现药物的组合使用情况,对其进行挖掘分析,发掘其治疗思想,可以直接指导疾病治疗。

从数据挖掘结果可以看出,运用最大频繁关联模式挖掘的方法进行处理分析,可以从不同的角度对冠心病血瘀证的诊断作出判断,其结果也能够一定程度满足临床的需求。本次研究发现,在所整理的百余例医案中,血瘀证中较为常见的证型为阳虚血瘀证和气滞血瘀证,桃仁、红花、当归、川芎是常用的四味药,胸痛、头痛、胸

闷是常见的三个症状,由于本次研究的样本量有限,处方量有限,患者的耐受性和依从性不同,尚不能达到完美的诊断价值,如何进一步将这些结果更好地结合在一起,得出一个更好的综合性结论,为临床制定满足不同需要的冠心病血瘀证证候诊断标准,尚需进一步按照临床综合评价的思路及临床设计进行探讨。

参 考 文 献

[1] 付长庚,高铸桦,王培利,等.冠心病血瘀证诊断标准研究[J].中国中西医结合杂志,2012,32(9):1285-1286.
[2] 陶涛淇.缺血性心脏病的命名及诊断标准[J].中华心血管病杂志,1981,9(1):75-76.
[3] 国家中医药管理局.中医病证诊断疗效标准[M].南京:南京大学出版社,1994:19.
[4] 李杰,袁肇凯.冠心病血瘀证的病机研究[J].中西医结合心脑血管病杂志,2006,4(7):613-615.
[5] 刘杰,葛晓玢,姚璐,等.基于矩阵的最大频繁项目集挖掘算法研究[J].电脑知识与技术,2011,7(29):7234-7236.
[6] 李超,余昭平.基于最大模式的关联规则挖掘算法研究[J].微计算机信息,2006,22(6):164-165,186.
[7] 吴毅挺.DartSpora数据挖掘平台的构建[J].浙江大学计算机科学与技术学院学报,2008,12(3):46-52.

(收稿:2013-10-11 修回:2016-12-31)

责任编辑:赵芳芳
英文责编:张晶晶