

• 循证医学与临床用药 •

基于 R 语言数据挖掘的中药治疗紧张型头痛的用药规律探析

樊雪鸣, 王柳丁, 申 伟*, 张允岭*

中国中医科学院西苑医院, 北京 100091

摘要:目的 基于 R 语言数据挖掘平台对现代文献中紧张型头痛(tension-type headache, TTH)的方药信息进行数据挖掘, 为 TTH 的中药治疗提供新思路。方法 计算机检索中国知网、万方、维普、中国生物医学、PubMed、Embase 和 Cochrane Library 共 7 个国内外数据库, 筛选中药治疗 TTH 的相关文献。采用 Excel 软件对方药信息进行提取, 采用 R 语言对中药使用频次、相关性、关联度和聚类等方面进行挖掘分析。结果 共提取出中药治疗 TTH 的 306 首经验方药, 中药共 238 味, 频次共 3187 次, 其中使用频次>100 次的药物依次为川芎、白芍、柴胡、甘草、当归和白芷, 使用频次>30 次的药物以解表药、活血化瘀药、清热药、补虚药、平肝熄风药为主, 相关系数较高的前 20 味药对组合多为补虚药和活血化瘀药, 关联分析和聚类分析形成以川芎、白芍、柴胡和当归为基础的核心药物组方。结论 中药治疗 TTH 以川芎、白芍、柴胡和当归为核心药物组方, 在此基础上加入针对风、火、痰、虚、瘀等致病因素的相应中药, 临床可取得较好疗效。建议后续研究应更多关注 TTH 的核心药物组方, 开展高质量的随机对照试验, 为 TTH 的临床治疗提供新思路。

关键词: 紧张型头痛; 数据挖掘; R 语言; 中药; 用药规律; 川芎; 白芍; 柴胡; 当归

中图分类号: R285 **文献标志码:** A **文章编号:** 0253-2670(2021)15-4614-12

DOI: 10.7501/j.issn.0253-2670.2021.15.020

Analysis of medication rules of traditional Chinese medicine in treatment of tension-type headache based on R language data mining

FAN Xue-ming, WANG Liu-ding, SHEN Wei, ZHANG Yun-ling

Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing 100091, China

Abstract: Objective To mine the prescription information of tension-type headache (TTH) in modern literature based on R language data mining platform, so as to provide new ideas for traditional Chinese medicine in the treatment of TTH. **Methods** Seven databases including CNKI, Wanfang, VIP, CBM, PubMed, EMBASE and Cochrane library were searched by computer to screen the related literatures. Excel software was used to extract prescription information, and the R language was used to mine and analyze the frequency, correlation, association and clustering of traditional Chinese medicine. **Results** A total of 306 empirical prescriptions for TTH were extracted. There were 238 Chinese herbs with a total frequency of 3187 times. Among them, Chuanxiong (*Chuanxiong Rhizoma*), Baishao (*Paeoniae Radix Alba*), Chaihu (*Bupleuri Radix*), Gancao (*Glycyrrhizae Radix et Rhizoma*), Danggui (*Angelicae Sinensis Radix*) and Baizhi (*Angelicae Dahuricae Radix*) were used more frequently than 100 times. The drugs with a frequency of more than 30 times were mainly for relieving exterior symptoms, promoting blood circulation and removing blood stasis, clearing heat, tonifying deficiency, calming liver and wind. The top 20 drug pairs with higher correlation coefficient were with efficacy for tonifying deficiency, promoting blood circulation and removing blood stasis. Through association and clustering analysis, the core prescriptions based on *Chuanxiong Rhizoma*, *Paeoniae Radix Alba*, *Bupleuri Radix* and *Angelicae Sinensis Radix* were formed. **Conclusion** *Chuanxiong Rhizoma*, *Paeoniae Radix Alba*, *Bupleuri Radix* and *Angelicae Sinensis Radix* are the core herbs of traditional Chinese medicine in the treatment of TTH. On this basis, the corresponding medicine for wind, fire, phlegm, deficiency, blood stasis and other pathogenic factors can be added to achieve better clinical efficacy. It is

收稿日期: 2021-02-16

基金项目: 国家中医药管理局中医药传承与创新“百千万”人才工程: 岐黄学者项目-国家中医药领军人才支持计划(国中医药人教发[2018]12号); 首都卫生发展科研专项(首发 2020-2-4173); 中央级公益性科研院所基本科研业务费专项(ZZ13-024-3)

作者简介: 樊雪鸣(1993—), 女, 在读博士, 研究方向为中医药防治脑病研究。Tel: 13439588985 E-mail: xueming_fan@126.com

***通信作者:** 申 伟, 博士, 住院医师, 从事中医药防治脑病研究。E-mail: 676665709@qq.com

张允岭, 主任医师, 博士生导师, 从事中医药防治脑病研究。E-mail: yunlingzhang2004@163.com

suggested that follow-up studies should pay more attention to the core drug formulation of TTH and carry out high-quality randomized controlled trials to provide new ideas for the clinical treatment of TTH.

Key words: tension-type headache; data mining; R language; traditional Chinese medicine; medication rules; *Chuanxiong Rhizoma*; *Paeoniae Radix Alba*; *Bupleuri Radix*; *Angelicae Sinensis Radix*

紧张型头痛 (tension-type headache, TTH) 是临床常见且反复发作的原发性头痛^[1-2], 患病率在一般人群中为 30%~78%^[3-4], 主要表现为双侧头部的压迫或紧张感, 常为轻度至中度疼痛^[5]。TTH 对中青年人群的情绪、睡眠等影响极大, 长时间严重且频繁的头痛不仅降低患者的生活质量, 且给社会带来巨大的经济损失^[1,6]。TTH 的发病机制目前尚不明确, 现代研究认为其发病主要与周围性疼痛、中枢敏化、神经心理、遗传、睡眠和一氧化氮等多种因素相关^[7]。TTH 的临床治疗以非甾体类抗炎药和抗抑郁药为主^[8], 但其口干、嗜睡、头晕和胃肠道不适等不良反应常不被患者所接受^[2]。

TTH 属于中医“头痛”“头风”范畴, 其病因复杂, 机制繁多, 致病因素主要涉及风、火、痰、瘀、虚 5 类^[9]。中医药在治疗 TTH 时, 常以行气解郁为大法, 针对不同病理因素配伍不同属性药物, 可有效缓解头痛症状, 预防复发, 且不良反应少, 胃肠道刺激小, 患者易于接受^[10-12]。然而, 由于多数临床研究仍停留在小样本随机对照试验和个案报告中^[13], 虽然中药疗效确切, 但涉及方药众多, 缺少对方药信息的整体把握。

中医方药在中药的配伍和方证对应等方面充分体现了中医辨证论治的特点, 利用合适的方法对方药信息进行研究, 可以发现临床用药规律, 解决新药研发等实际问题^[14]。数据挖掘是从大量数据中提取隐含的、人们事前未知的但又有潜在价值和特殊关系的信息和知识的过程^[14-15]。利用数据挖掘方法分析中医经验诊疗信息, 可实现中医方药经验的有效总结与传承创新^[16]。赵艳青等^[17]基于中医传承辅助平台系统对 TTH 的组方用药进行总结, 为中医药临床治疗提供了相应依据。但该文仅检索中国知网数据库, 纳入文献存在一定的偏倚, 且将提取的全部中药纳入聚类分析, 未考虑药物使用频次的影响, 此外, 中医传承辅助平台系统运用算法较为固定, 需要更加专业的编程分析软件针对不同数据类型设计个性化研究方案。

R 语言是世界公认开源编程软件, 通过安装相应程序包可实现对数据库数据的多种个性化分析功

能^[18-19]。相较于传统数据挖掘软件, R 语言操作灵活, 可根据数据类型设定不同研究方法, 使研究过程更加贴合研究者不同的分析需求^[20], 且 R 语言具有强大的数据可视化技术, 可利用计算机的图像处理功能, 将数据转换为直观图表, 有助于对复杂信息的深入理解和沟通交流^[21]。

本研究通过收集现代文献中中药治疗 TTH 的方药信息, 运用 R 语言对中药治疗 TTH 的用药规律进行数据挖掘, 为 TTH 的中药治疗提供新思路。

1 资料与方法

1.1 资料来源

通过计算机检索中国知网、万方、维普、中国生物医学 4 大中文数据库及 PubMed、Embase、Cochrane Library 3 大英文数据库, 收集中药或中成药治疗 TTH 的文献, 所有纳入文献均为全文文献, 语种限定为中文和英文。

1.2 检索策略

通过主题词与自由词结合的方式对检索词进行限定, 以“紧张型头痛”“紧张性头痛”“头风”“首风”“中药”“中成药”“草药”“汤”“丸”“散”为检索词检索中文数据库; 以“Tension-Type Headache”“Stress Headache”“Tension Headache”“Traditional Chinese Medicine”“Herbal Medicine”等为检索词检索英文数据库。检索年限限定为建库至 2020 年 9 月。

1.3 文献选择标准

1.3.1 文献纳入标准 ①研究类型: 实验性研究、观察性研究和名老中医经验等含有治疗方药的文献; ②研究对象: 临床明确诊断的 TTH 患者, 年龄≥18 岁, 不限性别和种族; ③干预措施: 文献中以中药和中成药等作为主要治疗手段, 包括药物组成明确的汤剂、颗粒剂、丸剂、胶囊、冲剂和散剂等内治法, 用药疗程和剂量不限; ④同一方药不同文献, 选取发表时间最近者录入; ⑤同一文献不同证型方药, 根据辨证类型分别录入。

1.3.2 文献排除标准 ①合并其他类型头痛; ②以民族医药作为主要治疗手段; ③方药组成不明确; ④综述、指南、系统评价、理论探讨和动物实验研究等研究类型文献。

1.4 文献资料提取

2 名研究者 (FXM 和 WLD) 根据文献纳排标准独立筛选文献、提取信息, 遇到分歧时可组内讨论或由第 3 名研究者 (SW) 进行裁决。运用 NoteExpress (v3.3.0.8102) 软件对文献进行管理, 通过阅读文献题目、摘要和全文, 确定最终纳入文献数量。运用 Microsoft Excel (v16.0.13426.20332) 软件建立 TTH 数据库, 采用背对背输入和第 3 方独立核查的方式保证数据的准确录入, 数据库录入信息包括: 编号、题目、方药名称、中药名称等。

1.5 文献数据规范

本研究参照《中国药典》2020 年版^[22]和 2007 年版《中药学》^[23], 对纳入文献方药中的中药名称进行规范化处理, 将中药别名统一修改为规范名称。如将“山萸肉”统一规范为“山茱萸”及“全虫”统一规范为“全蝎”等, 保留因炮制方法或产地不同而药性差别较大的中药, 如“地黄与熟地黄”“牛膝与川牛膝”“甘草与炙甘草”等。

1.6 数据统计方法

本研究运用 R 语言 (v3.6.1) 对数据库数据进行统计分析。运用 reshape 函数将规范后的中药进行二分类量化处理, 以方药名称为行名, 不同中药品种为列名, 对中药使用情况进行赋值, 使用赋值为 1, 未使用赋值为 0, 建立数据整形数据库。运用 ggplot2 函数对数据进行频次分析, 提取使用次数较多的中药绘制条形图, 直观反映 TTH 治疗

的有效中药品种。运用 corrplot 函数对数据进行相关分析, 提取临床治疗有效药对组合。运用 arules、arulesViz 函数进行关联分析, 对中药使用关联度进行可视化展示。运用 K-means、PAM 和 GMM 算法对数据进行聚类分析, 选取最佳聚类数目, 并分析每类中药功效特点, 直观反映中医方药组方规律。

2 结果

2.1 纳入文献情况

初步检索文献共计 13 005 篇, 其中中文文献 12 017 篇, 英文文献 988 篇。根据文献题目和作者进行查重, 剔除重复文献 5504 篇, 之后阅读文献题目和摘要初筛, 剔除不相关文献 6249 篇, 其他语种文献 40 篇, 剩余 1212 篇文献与 TTH 相关, 进一步阅读文献全文复筛, 剔除不符合纳入标准的文献 681 篇, 共计 531 篇文献纳入研究, 对文献方药信息进行提取, 剔除药物使用重复和无明确方药信息的文献, 最终得到经验方药共计 306 首。具体文献筛选流程见图 1。

2.2 频次分析

对 306 首经验方药进行中药频次分析, 共包含 238 味中药, 总使用频次 3187 次。其中使用频次 > 100 次的中药共 6 味, 依次为川芎 (221 次)、白芍 (181 次)、柴胡 (138 次)、甘草 (134 次)、当归 (127 次)、白芷 (110 次), 另有 152 味中药使用频次 ≤ 5 次, 占全部中药味数的 63.87%。见表 1。

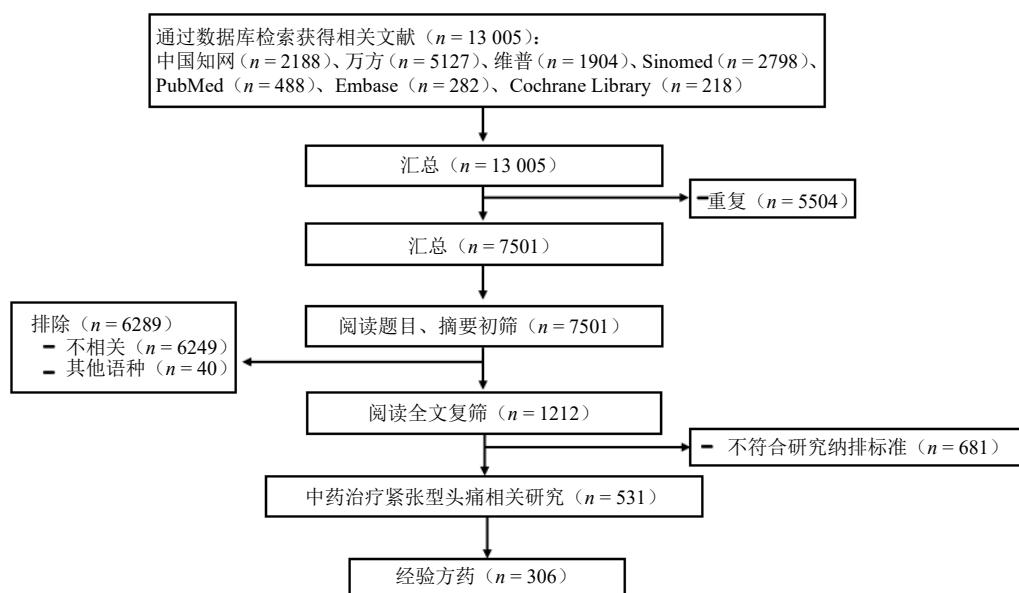


图 1 文献筛选流程

Fig. 1 Literature screening process

表 1 中药使用频次分析

Table 1 Frequency analysis of traditional Chinese medicine

中药使用频次 (n)	中药味数	占比/%
$n > 100$	6	2.52
$50 < n \leq 100$	6	2.52
$30 < n \leq 50$	19	7.98
$20 < n \leq 30$	14	5.88
$10 < n \leq 20$	20	8.41
$5 < n \leq 10$	21	8.82
$n \leq 5$	152	63.87
总计	238	100

提取使用频次 >30 次的 31 味中药绘制条形图, 并根据 2007 年版《中药学》^[23]对中药进行功效分类, 结果显示解表药 8 味: 柴胡、白芷、葛根、羌活、细辛、蔓荆子、菊花、薄荷; 活血化瘀药 6 味: 川芎、红花、桃仁、郁金、丹参、延胡索; 清热药 5 味: 甘草、地黄、黄芩、赤芍、栀子; 补虚药 4 味: 白芍、当归、白术、炙甘草; 平肝熄风药 4 味: 天麻、全蝎、钩藤、僵蚕; 理气药 2 味: 香附、陈皮; 利水渗湿药 1 味: 茯苓; 安神药 1 味: 酸枣仁。见图 2、3。

2.3 相关分析

相关分析可以反映数据间的相关性, 通常运用

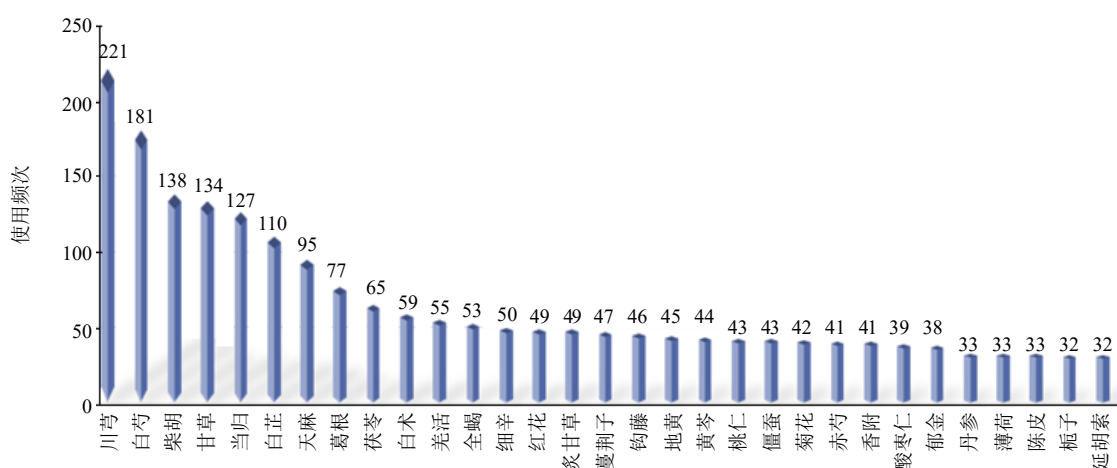


图 2 使用频次 >30 次的中药

Fig. 2 Traditional Chinese medicine used more than 30 times

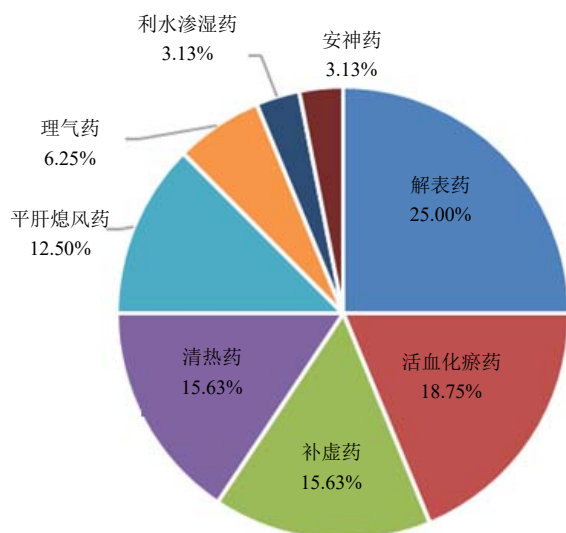


图 3 使用频次 >30 次的中药功效分类占比

Fig. 3 Proportion of efficacy classification of traditional Chinese medicine used more than 30 times

相关系数对相关性大小进行描述^[24]。通过对中药频次进行分析, 发现多数中药使用次数较少, 药物间相关性较低, 因此本研究仅提取使用频次 >5 次的中药进行相关分析, 包括中药 86 味, 总使用频次 2918 次。对 86 味中药进行 Pearson 相关系数分析, 运用 corrplot 函数对相关系数进行可视化展示, 见图 4。

根据药物间 Pearson 相关系数大小对药对组合进行排列, 正值表示两药间呈正相关, 即同增同减, 变化趋势相同; 负值表示两药间呈负相关, 变化趋势相反。提取相关系数较高的前 20 味药对组合, 见表 2。

2.4 关联分析

数据关联可以反映数据间的关联性, 通过分析这种关联性, 设置相关规则, 可以有效挖掘数据间隐含的规律^[25]。衡量规则强度的指标包括支持度、置信度和提升度 3 类。支持度表示 A 组药和 B 组药

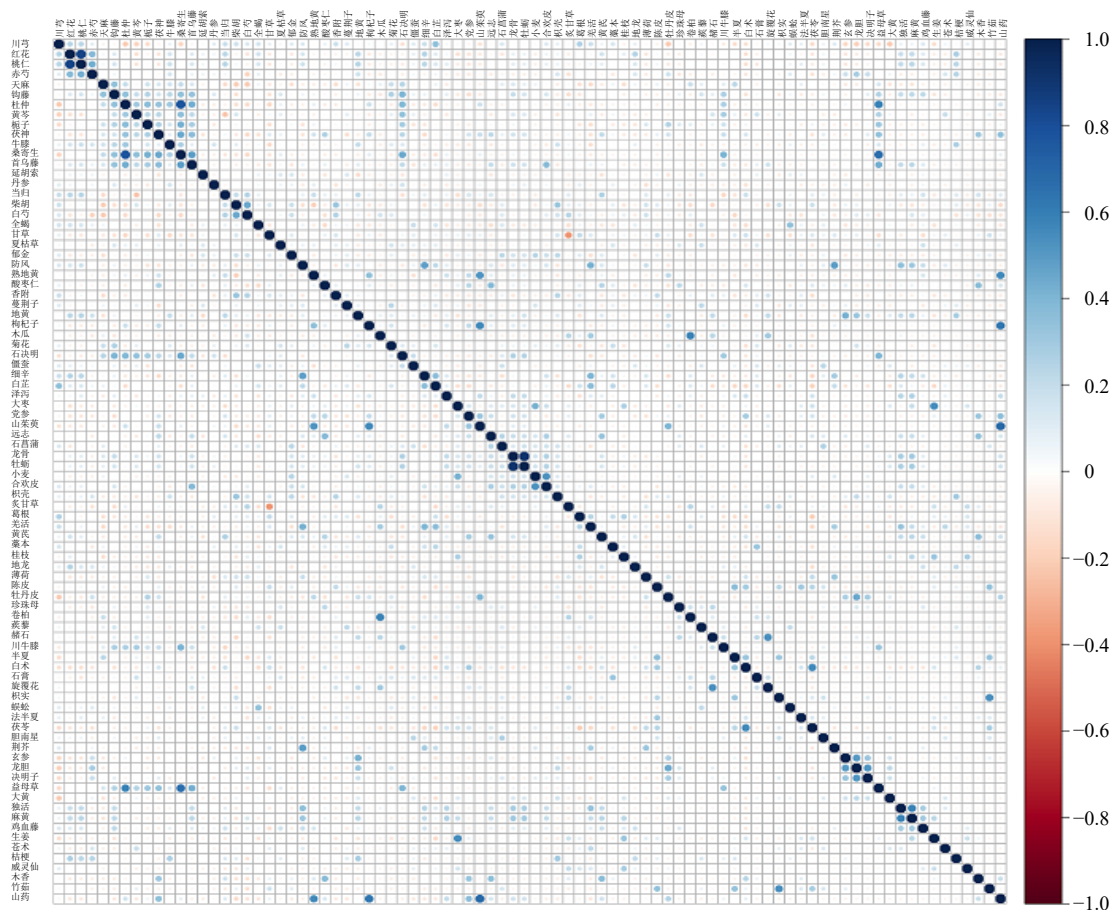


图 4 使用频次>5 次的中药 Pearson 相关系数图

Fig. 4 Pearson correlation coefficient of traditional Chinese medicine used more than five times

表 2 使用频次>5 次的中药相关系数前 20 味药对组合

Table 2 Correlation coefficients of top 20 drug pairs used more than five times

序号	药对组合	相关系数	序号	药对组合	相关系数
1	牡蛎-龙骨	0.92	11	茯苓-白术	0.56
2	桃仁-红花	0.85	12	山药-熟地黄	0.55
3	桑寄生-杜仲	0.78	13	生姜-大枣	0.54
4	山药-山茱萸	0.68	14	旋覆花-赭石	0.53
5	益母草-桑寄生	0.66	15	竹茹-枳实	0.53
6	山药-枸杞子	0.63	16	合欢皮-小麦	0.52
7	益母草-杜仲	0.59	17	山茱萸-熟地黄	0.52
8	麻黄-独活	0.58	18	决明子-龙胆	0.52
9	卷柏-木瓜	0.57	19	龙胆-玄参	0.50
10	山茱萸-枸杞子	0.56	20	荆芥-防风	0.48

同时使用占有所有药物使用的概率；置信度又称条件概率，表示使用 A 组药的同时使用 B 组药的概率；提升度表示 A 组药使用对 B 组药使用的提升作用，

提升度>1 表示规则为有效规则^[26]。

对使用频次>5 次的中药进行关联分析，设置最小支持度为 0.1，最小置信度为 0.8，共得到 59

条关联规则，提升度均 >1 ，为有效规则。其中 2 种药物关联规则 22 条，3 种药物关联规则 30 条，4 种药物关联规则 7 条，根据支持度大小对规则进行排列。见表 3~5。

运用 R 语言 `arules`、`arulesViz` 函数对 59 条关联规则进行可视化展示，气泡大小表示支持度，颜色深浅表示提升度，箭头指向次数多的中药提示为中心药物，在关联规则中使用次数较多。见图 5。

表 3 使用频次 >5 次的中药 2 阶关联规则

Table 3 Association rules of two kinds of traditional Chinese medicine used more than five times

序号	药对组合	支持度/%	置信度/%	序号	药对组合	支持度/%	置信度/%
1	{柴胡} $\Rightarrow$ {白芍}	37.58	83.33	12	{桃仁} $\Rightarrow$ {红花}	13.07	93.02
2	{柴胡} $\Rightarrow$ {川芎}	36.27	80.43	13	{红花} $\Rightarrow$ {桃仁}	13.07	81.63
3	{当归} $\Rightarrow$ {川芎}	33.99	81.89	14	{蔓荆子} $\Rightarrow$ {川芎}	12.75	82.98
4	{白芷} $\Rightarrow$ {川芎}	33.33	92.73	15	{香附} $\Rightarrow$ {川芎}	12.42	92.68
5	{葛根} $\Rightarrow$ {川芎}	23.20	92.21	16	{香附} $\Rightarrow$ {白芍}	12.09	90.24
6	{羌活} $\Rightarrow$ {川芎}	16.01	89.09	17	{僵蚕} $\Rightarrow$ {川芎}	11.76	83.72
7	{红花} $\Rightarrow$ {川芎}	15.36	95.92	18	{香附} $\Rightarrow$ {柴胡}	11.44	85.37
8	{全蝎} $\Rightarrow$ {川芎}	15.03	86.79	19	{赤芍} $\Rightarrow$ {川芎}	11.11	82.93
9	{细辛} $\Rightarrow$ {川芎}	14.71	90.00	20	{菊花} $\Rightarrow$ {川芎}	11.11	80.95
10	{桃仁} $\Rightarrow$ {川芎}	13.73	97.67	21	{郁金} $\Rightarrow$ {川芎}	10.46	84.21
11	{炙甘草} $\Rightarrow$ {白芍}	13.73	85.71	22	{郁金} $\Rightarrow$ {白芍}	10.13	81.58

表 4 使用频次 >5 次的中药 3 阶关联规则

Table 4 Association rules of three kinds of traditional Chinese medicine used more than five times

序号	药对组合	支持度/%	置信度/%	序号	药对组合	支持度/%	置信度/%
1	{川芎, 柴胡} $\Rightarrow$ {白芍}	29.74	81.98	16	{白芷, 葛根} $\Rightarrow$ {川芎}	11.76	94.74
2	{当归, 柴胡} $\Rightarrow$ {白芍}	20.59	87.50	17	{白芷, 羌活} $\Rightarrow$ {川芎}	11.76	90.00
3	{白芍, 白芷} $\Rightarrow$ {川芎}	19.93	92.42	18	{白芍, 香附} $\Rightarrow$ {川芎}	11.11	91.89
4	{柴胡, 甘草} $\Rightarrow$ {白芍}	19.28	83.10	19	{川芎, 香附} $\Rightarrow$ {白芍}	11.11	89.47
5	{柴胡, 白芷} $\Rightarrow$ {川芎}	16.67	94.44	20	{天麻, 白芷} $\Rightarrow$ {川芎}	11.11	89.47
6	{柴胡, 白芷} $\Rightarrow$ {白芍}	16.01	90.74	21	{柴胡, 葛根} $\Rightarrow$ {白芍}	11.11	85.00
7	{白芍, 葛根} $\Rightarrow$ {川芎}	15.69	92.31	22	{细辛, 白芷} $\Rightarrow$ {川芎}	10.78	89.19
8	{甘草, 白芷} $\Rightarrow$ {川芎}	15.69	92.31	23	{红花, 当归} $\Rightarrow$ {川芎}	10.46	96.97
9	{当归, 白芷} $\Rightarrow$ {川芎}	15.36	100.00	24	{柴胡, 香附} $\Rightarrow$ {白芍}	10.46	91.43
10	{当归, 甘草} $\Rightarrow$ {川芎}	14.38	81.48	25	{柴胡, 香附} $\Rightarrow$ {川芎}	10.46	91.43
11	{甘草, 葛根} $\Rightarrow$ {川芎}	13.73	89.36	26	{白芍, 香附} $\Rightarrow$ {柴胡}	10.46	86.49
12	{红花, 桃仁} $\Rightarrow$ {川芎}	12.75	97.50	27	{川芎, 香附} $\Rightarrow$ {柴胡}	10.46	84.21
13	{川芎, 桃仁} $\Rightarrow$ {红花}	12.75	92.86	28	{白芍, 茯苓} $\Rightarrow$ {柴胡}	10.46	82.05
14	{川芎, 红花} $\Rightarrow$ {桃仁}	12.75	82.98	29	{白芍, 羌活} $\Rightarrow$ {川芎}	10.13	88.57
15	{柴胡, 葛根} $\Rightarrow$ {川芎}	12.42	95.00	30	{川芎, 炙甘草} $\Rightarrow$ {白芍}	10.13	86.11

表 5 使用频次 >5 次的中药 4 阶关联规则

Table 5 Association rules of four kinds of traditional Chinese medicine used more than five times

序号	药对组合	支持度/%	置信度/%	序号	药对组合	支持度/%	置信度/%
1	{川芎, 当归, 柴胡} $\Rightarrow$ {白芍}	15.69	84.21	5	{当归, 白芍, 白芷} $\Rightarrow$ {川芎}	10.46	100.00
2	{柴胡, 白芍, 白芷} $\Rightarrow$ {川芎}	15.36	95.92	6	{柴胡, 白芍, 葛根} $\Rightarrow$ {川芎}	10.46	94.12
3	{川芎, 柴胡, 白芷} $\Rightarrow$ {白芍}	15.36	92.16	7	{川芎, 柴胡, 葛根} $\Rightarrow$ {白芍}	10.46	84.21
4	{川芎, 柴胡, 甘草} $\Rightarrow$ {白芍}	14.71	81.82				

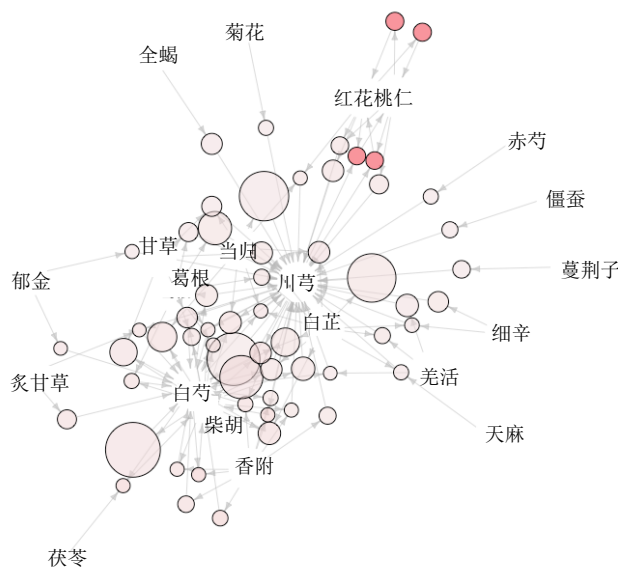


图 5 使用频次>5 次的中药关联规则分析

Fig. 5 Association analysis of traditional Chinese medicine used more than five time

2.5 聚类分析

聚类分析是根据数据库中数据的特征，将大量未知标注的数据划分为多个不同簇集的过程，簇内数据间呈相关关系，不同簇集数据间则不相关，簇集间存在差异性^[27]。

本研究对使用频次>5 次的 86 味中药进行聚类分析，运用 K-means、PAM 和 GMM 算法确定最佳聚类数目，并通过树状图层次结构将结果进行展示。

2.5.1 K-means 算法 对 86 味中药进行基于 K-means 算法的最佳聚类数目测试，结果提示当聚类数目为 1 时聚类结果最佳，但分类为 1 类时不能反映不同药物间的差异性，故该法舍弃不用。见图 6。

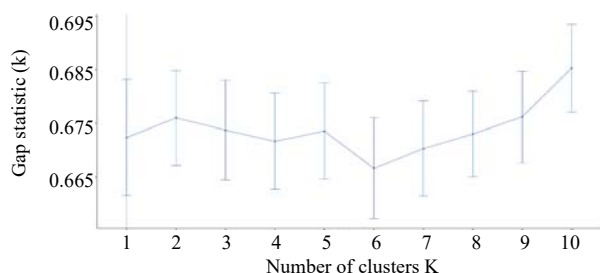


图 6 基于 K-means 算法确定最佳聚类数目

Fig. 6 Determining optimal number of clusters based on K-means algorithm

2.5.2 PAM 算法 K-means 算法以簇内点的均值作为簇的聚类中心，而由于孤立点的存在，致使聚类中心偏离数据的实际中心，影响算法准确率。PAM 算法则在聚类过程中同步过滤孤立点数据，避免单

独处理孤立点数据而增加的时间开销，提高了算法的准确率和执行效率^[28]。对 86 味中药进行基于 PAM 算法的最佳聚类数目测试，结果提示当聚类数目为 2 时聚类结果最佳。见图 7。

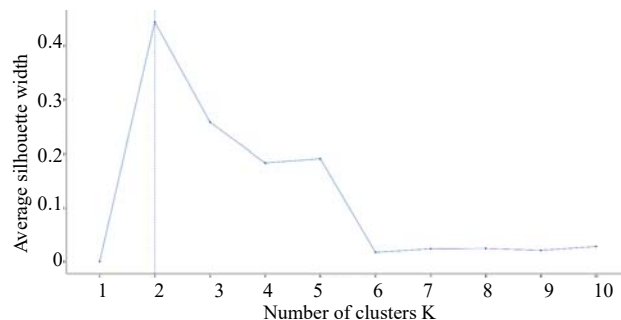


图 7 基于 PAM 算法确定最佳聚类数目

Fig. 7 Determining optimal number of clusters based on PAM algorithm

对 86 味中药进行层次聚类并绘制树状图，根据 PAM 算法将树状图分为药物特征相近的 2 类，结果提示第 1 类中药 10 味：川芎、天麻、当归、柴胡、白芍、甘草、白芷、葛根、白术、茯苓；第 2 类中药 76 味：红花、桃仁、赤芍、钩藤、杜仲、黄芩、栀子、茯神、牛膝、桑寄生、首乌藤、延胡索、丹参、全蝎、夏枯草、郁金、防风、熟地黄、酸枣仁、香附、蔓荆子、地黄、枸杞子、木瓜、菊花、石决明、僵蚕、细辛、泽泻、大枣、党参、山茱萸、远志、石菖蒲、龙骨、牡蛎、小麦、合欢皮、枳壳、炙甘草、羌活、黄芪、藁本、桂枝、地龙、薄荷、陈皮、牡丹皮、珍珠母、卷柏、蒺藜、赭石、川牛

膝、半夏、石膏、旋覆花、枳实、蜈蚣、法半夏、胆南星、荆芥、玄参、龙胆、决明子、益母草、大

黄、独活、麻黄、鸡血藤、生姜、苍术、桔梗、威灵仙、木香、竹茹、山药。见图 8。

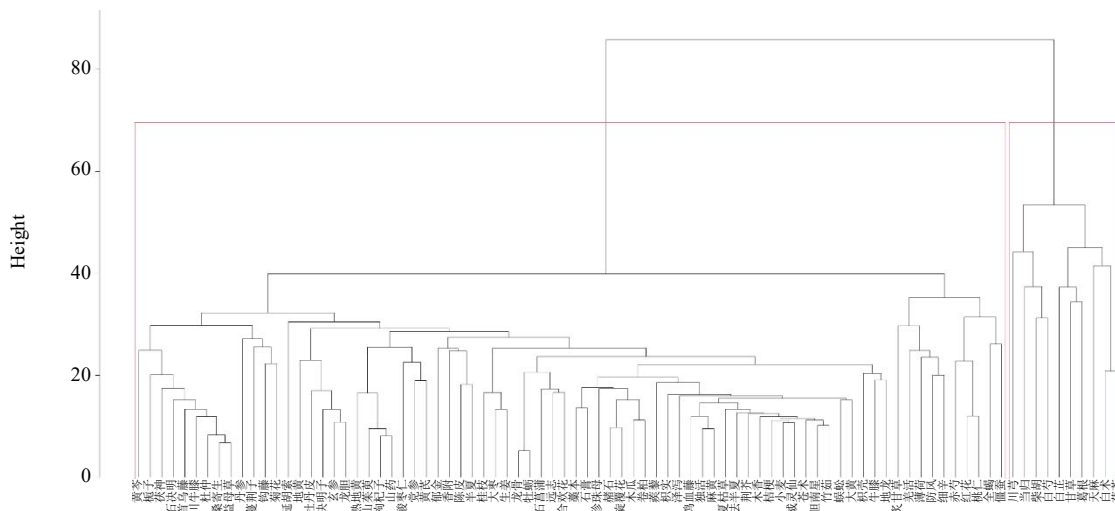


图 8 聚为 2 类的层次聚类树状图

Fig. 8 Hierarchical clustering tree diagram of clustering into two groups

2.5.3 GMM 算法 GMM 算法不仅可对数据进行簇集分类，同时可以计算数据属于某簇集的概率，与 K-means 算法相比，GMM 假设的模型分布与数据的真实分布一致性更强^[29]。运用 mclust 程序包对 86 味中药进行基于 GMM 算法的最佳聚类数目测试，结果提示共得到 4 种模型 1~9 组的分析结果，当聚类数目为 3 或 4 时聚类结果最佳。见图 9。

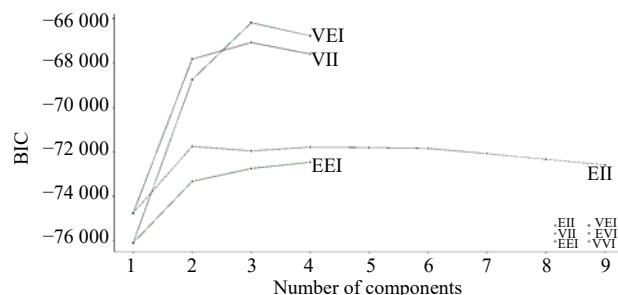


图 9 基于 GMM 算法确定最佳聚类数目

Fig. 9 Determining optimal number of clusters based on GMM algorithm

根据 GMM 算法将树状图分为药物特征相近的 3 类，结果提示第 1 类中药 4 味：川芎、当归、柴胡、白芍；第 2 类中药 6 味：天麻、甘草、白芷、葛根、白术、茯苓；第 3 类中药 76 味：红花、桃仁、赤芍、钩藤、杜仲、黄芩、栀子、茯神、牛膝、桑寄生、首乌藤、延胡索、丹参、全蝎、夏枯草、郁金、防风、熟地黄、酸枣仁、香附、蔓荆子、地黄、枸杞子、木瓜、菊花、石决明、僵蚕、细辛、泽泻、大枣、党参、山茱萸、远志、石菖蒲、龙骨、牡蛎、小麦、合欢皮、枳壳、炙甘草、羌活、黄芪、藁本、桂枝、地龙、薄荷、陈皮、牡丹皮、珍珠母、卷柏、蒺藜、赭石、川牛膝、半夏、石膏、旋覆花、枳实、蜈蚣、法半夏、胆南星、荆芥、玄参、龙胆、决明子、益母草、大黄、独活、麻黄、鸡血藤、生姜、苍术、桔梗、威灵仙、木香、竹茹、山药。见图 10。

大枣、党参、山茱萸、远志、石菖蒲、龙骨、牡蛎、小麦、合欢皮、枳壳、炙甘草、羌活、黄芪、藁本、桂枝、地龙、薄荷、陈皮、牡丹皮、珍珠母、卷柏、蒺藜、赭石、川牛膝、半夏、石膏、旋覆花、枳实、蜈蚣、法半夏、胆南星、荆芥、玄参、龙胆、决明子、益母草、大黄、独活、麻黄、鸡血藤、生姜、苍术、桔梗、威灵仙、木香、竹茹、山药。见图 10。

根据 GMM 算法将树状图分为药物特征相近的 4 类，结果提示第 1 类中药 4 味：川芎、当归、柴胡、白芍；第 2 类中药 3 味：白芷、甘草、葛根；第 3 类中药 3 味：天麻、白术、茯苓；第 4 类中药 76 味：红花、桃仁、赤芍、钩藤、杜仲、黄芩、栀子、茯神、牛膝、桑寄生、首乌藤、延胡索、丹参、全蝎、夏枯草、郁金、防风、熟地黄、酸枣仁、香附、蔓荆子、地黄、枸杞子、木瓜、菊花、石决明、僵蚕、细辛、泽泻、大枣、党参、山茱萸、远志、石菖蒲、龙骨、牡蛎、小麦、合欢皮、枳壳、炙甘草、羌活、黄芪、藁本、桂枝、地龙、薄荷、陈皮、牡丹皮、珍珠母、卷柏、蒺藜、赭石、川牛膝、半夏、石膏、旋覆花、枳实、蜈蚣、法半夏、胆南星、荆芥、玄参、龙胆、决明子、益母草、大黄、独活、麻黄、鸡血藤、生姜、苍术、桔梗、威灵仙、木香、竹茹、山药。见图 11。

3 讨论

本研究利用 R 语言数据挖掘平台，对收集到的 306 首中药治疗 TTH 的方药信息进行个性化编程，

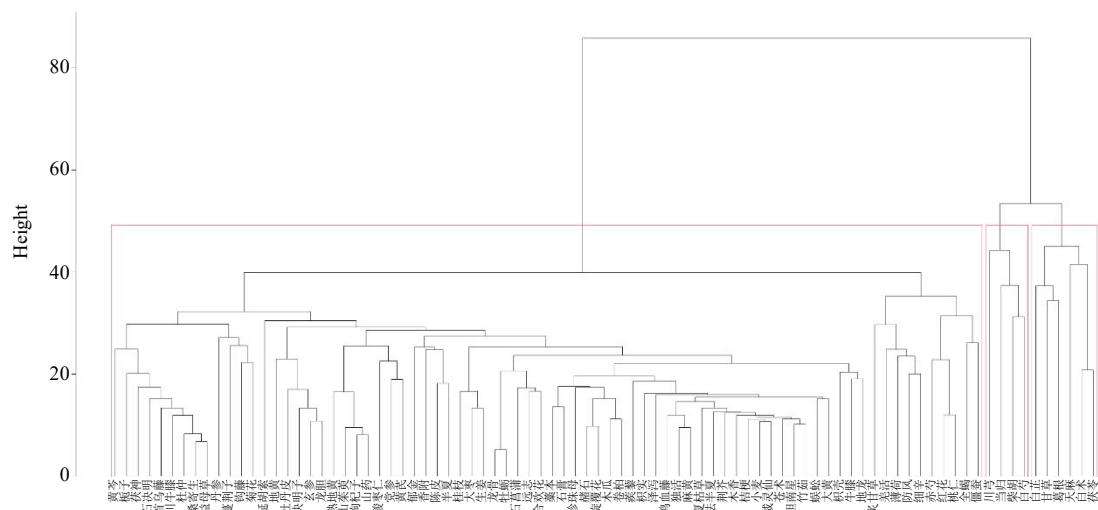


图 10 聚为 3 类的层次聚类树状图

Fig. 10 Hierarchical clustering tree diagram of clustering into three groups

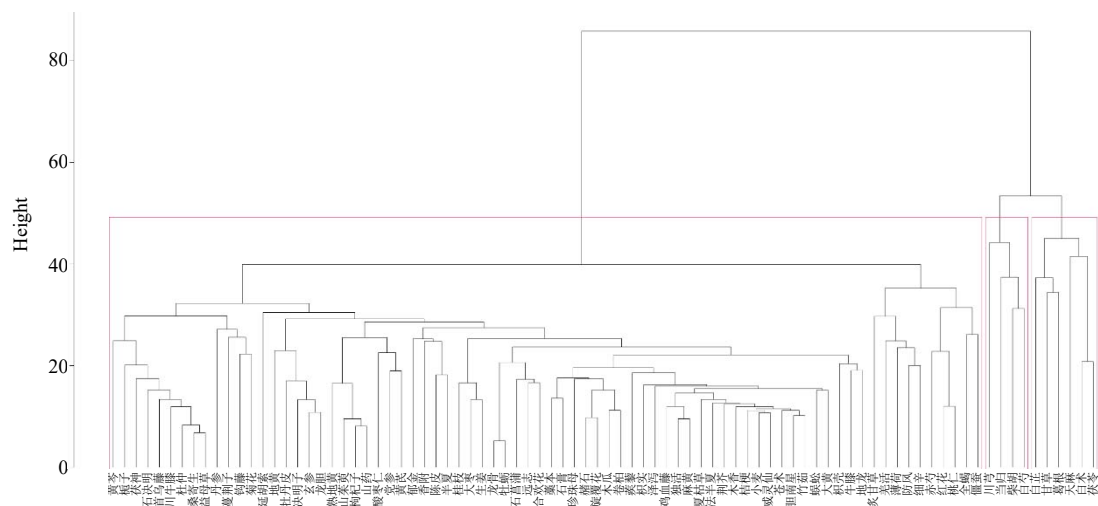


图 11 聚为 4 类的层次聚类树状图

Fig. 11 Hierarchical clustering tree diagram of clustering into four groups

深入挖掘分析 TTH 的用药规律,得出临床治疗 TTH 的核心药物组方。

3.1 结果分析

3.1.1 频次分析 对 306 首经验方药进行中药频次分析,共得到中药 238 味,总使用频次 3187 次。其中使用频次较多的药物依次为川芎、白芍、柴胡、甘草、当归和白芷,与既往数据挖掘的研究相比,药物使用频次结果基本一致^[17]。针对不同使用频次进行相应分组,结果显示多数中药使用频次不足 5 次,占全部中药味数的 63.87%,提示目前临床上对于中药治疗 TTH 药物运用繁杂,具体治疗方案未形成统一规范。对使用次数较多的 31 味中药进行功效分析,发现以解表药、活血化瘀药、清热药、

补虚药、平肝熄风药为主,与目前 TTH 从风、火、痰、虚、瘀等方面进行治疗的思想基本吻合^[10]。

3.1.2 相关分析 对使用次数较多的 86 味中药进行相关分析,结果显示相关系数较高的前 20 味药对组合多为补虚药和活血化瘀药,且多为牡蛎-龙骨、桃仁-红花、桑寄生-杜仲等临床常见药对组合。由于本研究在进行相关分析仅考虑 Pearson 相关系数,即仅对药物之间的相关性进行总结,而具体药物使用频次则未纳入研究,如牡蛎和龙骨使用频次分别为 20 次和 19 次,虽然两药相关性最高,临床上常作为药对使用,但在 TTH 的药物使用中其运用次数较少,相关系数的研究结果不能反映中药治疗 TTH 的特异性药对组合,故本研究相关分析结

果仅作为参考。

3.1.3 关联分析 对使用频次>5 次的 86 味中药进行关联分析, 结果显示{柴胡} $\Rightarrow$ {白芍}、{川芎, 柴胡} $\Rightarrow$ {白芍}和{川芎, 当归, 柴胡} $\Rightarrow$ {白芍}分别为 2 种、3 种和 4 种药物关联规则中支持度最高的组合, 气泡图中川芎、柴胡和白芍为箭头指向次数最多的药物, 2 项研究均提示川芎、白芍、柴胡和当归为中药治疗 TTH 的核心药物组方。中医认为, 川芎辛温, 活血行气、祛风止痛; 当归甘辛温, 补血活血、化瘀止痛, 两药配伍, 气血兼顾, 有补血活血之功; 白芍苦酸微寒, 养血柔肝、平抑肝阳; 柴胡辛苦微寒, 疏肝解郁、祛达少阳风邪, 两药配伍, 疏肝敛阴, 有调理肝胆气机之效。现代药理学研究证实, 川芎中含有多达 45 种苯酞类化合物^[30], Z-藁本内酯是其中主要的活性成分, 具有扩张血管、抗血小板聚集、镇痛等调节心脑血管和神经系统的作用^[31]; 当归中含有的多糖能够促进造血细胞的分化和增殖, 刺激造血生长因子的释放, 从而改善人体的造血功能^[32], 两药配伍能明显改善血瘀大鼠血清中血管活性物质和黏附分子表达, 具有较强的活血化瘀的功效^[33]。白芍的药效成分主要为苷类物质, 其中芍药苷占总苷量的 90% 以上, 是白芍的主要有效成分^[34], 白芍总苷对痛觉有一定的影响, 其对白细胞介素-2 (interleukin-2, IL-2) 的抑制作用可能是其镇痛的主要机制^[35]; 柴胡的化学成分主要为三萜皂苷类化合物柴胡皂苷, 通过影响相关疼痛因子的释放, 降低患者的疼痛等级, 达到镇痛的作用^[36], 两药配伍, 柴胡可以促进白芍中芍药苷的溶出, 增强养血柔肝的功效^[37]。由此可见, 4 药合用, 可达行气解郁、补血活血、祛风止痛之功。

3.1.4 聚类分析 对使用频次>5 次的 86 味中药进行聚类分析, 结果显示分为 2 类、3 类和 4 类的层次聚类树状图中仅有 10 味中药 (川芎、天麻、当归、柴胡、白芍、甘草、白芷、葛根、白术、茯苓) 改变簇集关系, 另有 76 味中药簇集分类始终相同。76 味中药中使用频次最高的为羌活 (55 次)、全蝎 (53 次) 和细辛 (50 次), 占 306 首方药使用频次的 16%~18%, 药物使用次数较少, 由于多数中药的使用与医家根据患者不同症状辨证论治加减相关, 因此考虑该类 76 味中药与 TTH 的临床治疗相关性较小, 暂不纳入 TTH 的核心中药处方。改变簇集关系的 10 味中药与使用频次较高的前 10 味中药名称相同, 且第 1 类 4 味中药符合关联分析中核

心药物组方, 因此认为该 4 味中药为治疗 TTH 的中心药物; 第 2 类中药为白芷、甘草、葛根, 与解表清热之柴葛解肌汤的药物组成有相似之处, 对于以风邪为主的外感头痛或可有较好疗效; 第 3 类中药为天麻、白术、茯苓, 与化痰熄风之半夏白术天麻汤的药物组成相似, 可治疗痰浊上蒙之头痛。由此分析, 考虑中药治疗 TTH 的中心药物组方主要为川芎、白芍、柴胡和当归, 在 4 味中药的基础上加入针对风、火、痰、虚、瘀等致病因素的相应中药, 对于 TTH 的治疗或可取得较好疗效。

3.2 研究局限性

由于以下几方面原因, 本研究运用 R 语言数据挖掘平台分析 TTH 用药规律存在一定局限性: ①本研究采用主题词与自由词相结合的方式检索 4 大中文和 3 大英文数据库, 与既往研究^[17]相比扩大检索范围, 但仍缺乏临床注册网站的检索, 且未进行手工检索, 研究纳入 306 首方药信息, 整体数据偏少, 存在因纳入文献数量不足而影响研究结果的可能; ②本研究为增加纳入研究数量, 对检索到的病例系列和个案报告均进行了分析, 方药信息可能存在特异性; ③部分纳入研究未明确描述中医辨证分型, 因此本研究对中医辨证分型未进行统一规范, 符合 TTH 诊断的中医方药治疗均纳入研究, 但纳入研究证型各异, 证型诊断标准不同, 且部分研究缺乏证型信息, 可能存在因辨证分型不同而对研究结果产生影响的情况; ④考虑中药名称存在别称, 本研究参照《中国药典》2020 年版和 2007 年版《中药学》对名称进行规范, 但仍保留部分由于炮制方法或产地不同而药性差别较大的中药, 且部分文献中并未明确书写中药炮制或产地, 在前期数据处理中对该部分中药信息的保留对研究结果可能产生影响; ⑤本研究仅对中药药味进行数据挖掘, 未考虑中药相关剂量对方药组方的影响, 有待临床进一步研究; ⑥目前临床上运用 R 语言数据挖掘方法进行中药用药规律探索的相关文献较少, 且 R 语言分析方法缺乏统一规范, 部分医家认为聚类分析中最佳聚类数目的确定可通过树状图提示设定^[38-40], 另有部分医家认为可运用 K-means 等不同算法计算分类数目^[24,41], 本研究运用 PAM 和 GMM 算法对中药进行分类, 分类结果不同可能对研究结果产生一定影响。

3.3 展望

本研究表明中药治疗 TTH 的核心药物组方主

要为川芎、白芍、柴胡和当归，现代药理学研究也证实 4 药组合有较好的镇痛功效。提示中药在治疗 TTH 方面可以川芎、白芍、柴胡和当归为基础方，根据患者不同症状特点辨证论治，加入针对风、火、痰、虚、瘀等致病因素的相应中药，临床可取得较好疗效。

由于目前 TTH 的发病机制尚不明确，且口服化学药的不良反应临床较为常见，中药治疗 TTH 得到越来越多医家的关注。未来研究需更多关注 TTH 的核心药物组方，设计高质量、严格的临床研究方案，提高临床研究质量，深入探讨核心药物组方的临床有效性和安全性，为 TTH 的中药治疗和新药研发提供更多新思路。

4 结语

本研究表明中药治疗 TTH 的药物以解表药、活血化瘀药、补虚药和清热药为主，川芎、白芍、柴胡和当归为中药治疗的核心药物组方，在此基础上加入针对风、火、痰、虚、瘀等致病因素的相应中药，临床可取得较好疗效。后续研究仍需更多关注 TTH 的核心药物组方，开展高质量的随机对照试验，为 TTH 的临床治疗提供新思路。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 张虹玮, 白云, 孙亚萍, 等. 紧张型头痛的中医研究现状 [J]. 四川中医, 2020, 38(2): 216-219.
- [2] Headache classification committee of the international headache society (IHS) the international classification of headache disorders, 3rd edition [J]. *Cephalalgia*, 2018, 38(1): 1-211.
- [3] Wieckiewicz M, Grychowska N, Zietek M, et al. Evidence to use botulinum toxin injections in tension-type headache management: A systematic review [J]. *Toxins* (Basel), 2017, 9(11): E370.
- [4] Jackson J L, Mancuso J M, Nickoloff S, et al. Tricyclic and tetracyclic antidepressants for the prevention of frequent episodic or chronic tension-type headache in adults: A systematic review and meta-analysis [J]. *J Gen Intern Med*, 2017, 32(12): 1351-1358.
- [5] Linde K, Allais G, Brinkhaus B, et al. Acupuncture for the prevention of tension-type headache [J]. *Cochrane Database Syst Rev*, 2016, 4: CD007587.
- [6] Santesso N, Wieland L S. A summary of a cochrane review: Acupuncture for the prevention of tension-type headache [J]. *Eur J Integr Med*, 2016, 8(4): 324-325.
- [7] 刘豆豆, 雷琦, 杨谦, 等. 紧张型头痛发病机制的研究进展 [J]. 中国医刊, 2019, 54(1): 19-23.
- [8] Steiner T J, Jensen R, Katsarava Z, et al. Aids to management of headache disorders in primary care (2nd edition): On behalf of the European headache federation and lifting the burden: The global campaign against headache [J]. *J Headache Pain*, 2019, 20(1): 57.
- [9] 段文慧, 袁永, 宋军营, 等. 中医药治疗紧张型头痛的研究进展 [J]. 中国中医药现代远程教育, 2017, 15(5): 141-143.
- [10] 莫毅, 胡玉英. 中医药治疗紧张型头痛的研究进展 [J]. 大众科技, 2019, 21(4): 84-87.
- [11] 董宁. 中成药治疗紧张型头痛的 Meta 分析 [D]. 济南: 山东中医药大学, 2012.
- [12] 王凯. 紧张型头痛临床研究进展 [J]. 光明中医, 2020, 35(16): 2610-2614.
- [13] 龚帆, 俞晓飞. 慢性紧张型头痛中医药治疗新进展 [J]. 上海中医药杂志, 2017, 51(S1): 262-264.
- [14] 高毅超, 王凡, 郭晶. 数据挖掘技术在中医药领域中的应用概况 [J]. 湖南中医杂志, 2019, 35(7): 182-185.
- [15] 许雪莲, 吴昆仑. 数据挖掘技术在中医药研究中的应用 [J]. 河南中医, 2020, 40(11): 1633-1637.
- [16] 张吉英, 马金刚, 曹慧. 数据挖掘技术在中医药研究中的应用 [J]. 山东中医药大学学报, 2014, 38(1): 83-85.
- [17] 赵艳青, 滕晶. 基于中医传承辅助平台系统紧张型头痛的组方用药规律分析 [J]. 世界中医药, 2015, 10(8): 1261-1263.
- [18] 李杏, 刘涛, 肖建鹏, 等. R 语言在公共卫生领域的应用: Meta 分析 [J]. 华南预防医学, 2020, 46(5): 568-570.
- [19] 刘钰, 余卓芮, 刘岱宁. R 语言可视化的优势及其在空间统计教学中的应用 [J]. 高教论坛, 2020(5): 30-33.
- [20] 李一林, 蒋春波, 金伟民, 等. 基于 R 语言的中医治疗反复发作性泌尿系感染组方规律研究 [J]. 中医药信息, 2020, 37(1): 24-30.
- [21] 崔尧, 李建军, 丛芳, 等. 基于 R 语言的数据可视化技术在脊髓损伤康复中的应用 [J]. 中国数字医学, 2020, 15(10): 85-87.
- [22] 中国药典 [S]. 一部. 2020: 1-616.
- [23] 高学敏. 中药学 [M]. 北京: 中国中医药出版社, 2007: 51-523.
- [24] 林景峰, 胡文悦, 刘甘露, 等. 基于数据挖掘技术分析韩振蕴教授治疗抑郁症的临床用药规律 [J]. 中西医结合心脑血管病杂志, 2020, 18(10): 1531-1539.
- [25] 王琼, 杨明杰, 闫润珍. 大规模数据库中关联规则的发现与研究 [J]. 科技经济导刊, 2020, 28(34): 41-42.
- [26] 王燕, 周冰倩, 庞龙, 等. 基于数据挖掘探讨中医药治疗年龄相关性黄斑变性的用药规律 [J]. 中药药理与

- 临床, 2021, 32(4): 579-584.
- [27] Muhammad A. GMM 树和森林: 用于估计高维复杂数据中聚类数目的分层算法 [D]. 深圳: 深圳大学, 2020.
- [28] 卜旭松, 刘立波, 石磊. 基于 PAM 和簇阈值的改进 K-Means 聚类算法 [J]. 湖北工程学院学报, 2015, 35(3): 36-39.
- [29] 张鹏飞, 董敏周, 端军红. 基于集成 GMM 聚类的少标记样本图像分类 [J]. 西北工业大学学报, 2019, 37(3): 465-470.
- [30] 李海刚, 胡晒平, 周意, 等. 川芎主要药理活性成分药理研究进展 [J]. 中国临床药理学与治疗学, 2018, 23(11): 1302-1308.
- [31] 韩炜. 川芎的化学成分与药理作用研究进展 [J]. 中国现代中药, 2017, 19(9): 1341-1349.
- [32] 赵静, 夏晓培. 当归的化学成分及药理作用研究现状 [J]. 临床合理用药杂志, 2020, 13(6): 172-174.
- [33] 李伟霞, 王晓艳, 唐进法, 等. 基于 PLS-DA 和多指标综合指数法研究当归-川芎药对对急性血瘀大鼠血清中血管活性物质和黏附分子的影响 [J]. 药学报, 2019, 54(11): 1909-1917.
- [34] 杨丽梅, 岳广欣. 白芍总苷及其神经精神药理作用研究概况 [J]. 河北中医, 2017, 39(4): 614-618.
- [35] 李乃谦. 探讨白芍的药理作用及现代研究进展 [J]. 中医临床研究, 2017, 9(20): 137-138.
- [36] 辛国, 赵昕彤, 黄晓巍. 柴胡化学成分及药理作用研究进展 [J]. 吉林中医药, 2018, 38(10): 1196-1198.
- [37] 李春香, 丁里玉, 丁芳, 等. 白芍养血柔肝止痛作用与其配伍及芍药苷含量的相关性研究 [J]. 中药新药与临床药理, 2011, 22(1): 54-56.
- [38] 林明, 文豪, 李修身, 等. 基于数据挖掘的尿浊用药规律分析 [J]. 亚太传统医药, 2020, 16(3): 160-163.
- [39] 李一林, 蒋春波, 金伟民, 等. 基于 R 语言的中医治疗反复发作性泌尿系感染组方规律研究 [J]. 中医药信息, 2020, 37(1): 24-30.
- [40] 邹敏, 农志飞, 梅小平, 等. 基于数据挖掘防治小儿反复呼吸道感染方药配伍规律研究 [J]. 重庆医学, 2017, 46(35): 4980-4982.
- [41] 潘俊杰, 朱黎红, 宋康, 等. 基于 R 语言探索宋康教授运用中药辅助吉非替尼辨治肺腺癌的用药规律 [J]. 浙江中医药大学学报, 2020, 44(5): 419-425.

[责任编辑 潘明佳]