

基于《中国方剂数据库》挖掘便秘组方配伍规律

盛野¹ 吕晓东¹ 庞立健² 安阳¹ 王佳然¹ 梁元钰¹ 张思琪¹ 毛清雯¹

(1 辽宁中医药大学, 沈阳, 110847; 2 辽宁中医药大学附属医院, 沈阳, 110032)

摘要 目的:分析《中国方剂数据库》收录治疗便秘方剂组方用药配伍规律,为临床治疗便秘提供研究依据。方法:选取并筛选《中国方剂数据库》收录治疗便秘方剂,采用频次分析、因子分析、系统聚类、关联规则分析等数据挖掘方法对筛选方剂组方用药进行分析。结果:共筛选287首治疗便秘的方剂,涉及中药211味,出现频数最高的单味药为甘草(109次),其次为大黄(103次);因子分析可提取出25个主成分;聚类数目为5类时较好,常用药物有大黄、黄芩、当归、芍药、陈皮、木香、地黄、人参等,演化得到核心组合10个;关联分析显示:大黄配伍黄芩、栀子;大黄配伍芒硝;甘草配伍栀子、黄芩;大黄配伍黄连为常用药对。结论:《中国方剂数据库》收录治疗便秘方剂药类以补虚、解表、清热、理气、泻下以及化痰平喘为主。而大黄→槟榔;当归→地黄;甘草→陈皮;羌活→防风;芍药→当归值得进一步讨论。

关键词 数据挖掘;便秘;配伍规律;系统聚类分析;关联规则分析

Study on Compatibility Rules in Treating Constipation Basing on Mining TCM Formula Database

Sheng Ye¹, Lyu Xiaodong¹, Pang Lijian², An Yang¹, Wang Jiaran¹, Liang Yuanyu¹, Zhang Siqi¹, Mao Qingwen¹

(1 Liaoning University of Traditional Chinese Medicine, Shenyang 110847, China; 2 Affiliated Hospital of Liaoning University of Traditional Chinese Medicine, Shenyang 110032, China)

Abstract Objective: To analyze the compatibility rule of Chinese Prescription Database in the treatment of constipation, and to provide basis for clinical treatment of constipation. **Methods:** Based on the screening of formula in the treatment of constipation in the Chinese Prescription Database, we collected and analyzed the prescription by frequency analysis, factor analysis, system clustering, association rule analysis and other data mining methods. **Results:** A total of 287 prescriptions for constipation were screened, involving 211 Chinese herbs, and the most frequent one was Radix Glycyrrhizae (109 times), followed by Radix et Rhizoma Rhei (103 times). Factor analysis can extract 25 principal components. The cluster number of 5 kinds was good, and commonly used drugs were Radix et Rhizoma Rhei, Radix Scutellariae, Radix Angelicae Sinensis, Paeonia Lactiflora, Pericarpium Citri Reticulatae, Radix Aucklandiae, Radix Rehmanniae Recens, Radix Ginseng, etc., and 10 core combinations were obtained. The correlation analysis showed that Radix et Rhizoma Rhei combined with Radix Scutellariae and Fructus Gardeniae; Radix et Rhizoma Rhei with Natrii Sulfas; Radix Glycyrrhizae with Fructus Gardeniae and Radix Scutellariae; Radix et Rhizoma Rhei with Rhizoma Coptidis were known drug pair. **Conclusion:** The drugs of formula for constipation in Chinese Prescription Database were mainly tonifying deficiency, relieving superficies, clearing away heat, regulating qi, purging purgation and eliminating phlegm and relieving wheezing. And Radix et Rhizoma Rhei→ Semen Arcae; Radix Angelicae Sinensis→ Radix Rehmanniae Recens; Radix Glycyrrhizae→ Pericarpium Citri Reticulatae; Rhizoma et Radix Notopterygii→ Radix Saposchnikoviae; Paeonia Lactiflora→ Radix Angelicae Sinensis are worthy of further discussion, which will provide some reference for the research and development of new drugs of TCM for constipation.

Key Words Data mining; Constipation; Formulating rules; Hierarchical clustering; Association rules

中图分类号: R243; R574 文献标识码: A doi: 10.3969/j.issn.1673-7202.2019.08.054

便秘是指大肠传导功能失常,导致大便秘结,排便周期延长;或周期不长,但粪质干结,排便困难;或粪质不硬,虽有便意,但排出不畅的病证^[1]。《黄帝内经》首载便秘之症,称其为“后不利”“大便难”。汉·张仲景《伤寒杂病论》中将便秘称为“脾约”。刘完素首倡实秘、虚秘之别,《素问病机气宜保命

集·泻痢论》载:“凡脏腑之秘,不可一例治疗,有虚秘,有实秘,胃实而秘者,能饮食小便赤……胃虚而秘者,不能饮食,小便清利”。而便秘一名首见于清代沈金鳌的著作《杂病源流犀烛》中,并沿用至今。络脉从经脉分出后,逐级细分为别络至缠络、系络直至其末端孙络。随着络脉分支愈加细窄迂曲,络中

基金项目:国家中医药管理局中医络病重点学科建设项目(T0302);国家自然科学基金面上项目(81373579);国家自然科学基金青年基金项目(81403290);辽宁省科技厅计划项目(2012225018-202);辽宁省中临床重点学(专)科服务能力建设项目(2013-LNZYXZK-2);辽宁省高等学校科学研究一般项目(L2014369)

作者简介:盛野(1994.11—),男,硕士研究生,研究方向:中医药治疗肺系疾病的基础和临床研究,E-mail:windsor_van_saul@126.com

通信作者:吕晓东(1966.01—),女,博士,教授,研究方向:中医药治疗肺系疾病的基础和临床研究,E-mail:deanoftcm@126.com

气血的运行渐趋缓慢,一旦邪客络脉,络中气血的输布环流将受到影响,易致络脉瘀滞状态。气化功能失常,络气郁滞,或络气虚乏推运无力而滞,气化不利则津凝为痰,气失流畅则血行涩滞,痰瘀阻滞,络脉由滞到瘀,气血瘀塞不通。痰瘀阻络进一步影响气的升降出入运动而加重气滞。大肠传化功能的正常运行以及大便的通畅调合依赖于气机运行的调畅。例如肺主一身之气,与大肠相表里,若肺气失宣,壅滞不降,大肠传导阻滞则发为便秘。中药方剂是中医治疗的主要手段,基于博大而深厚的传统理论,其组成配伍规律庞杂多变。目前,基于数据挖掘技术对中医药方剂的研究主要集中于药证结合、中药组合、方剂配伍以及核心方药的发现方面。

基于络病理论,我们以国家人口与健康科学数据共享平台和中国中医科学院中医药信息研究所提供的《中国方剂数据库》为检索数据库,应用频数分析、因子分析^[2]、系统算法聚类分析^[3]以及 Apriori 算法关联规则分析^[4]等数据挖掘技术对治疗便秘方剂进行中药组方用药配伍规律分析,筛选所有便秘方剂中涉及的高频中药并从其药性、四气五味、归经配伍特点等方面加以探讨和研究,以期治疗便秘方剂的理论知识及临床治疗便秘的组方用药提供思路,进而提高临床治疗效果。

1 资料与方法

1.1 数据来源 选取国家人口与健康科学数据共享平台和中国中医科学院中医药信息研究所提供的《中国方剂数据库》中的方剂专题服务数据,包括方剂名称、药物组成、功效、主治。

1.2 数据筛选 以《中国方剂数据库》中的方剂专题服务数据为检索数据库,以方剂的主治、功效为依据,主治中明确载有脾约、闭、阴结、阳结、虚秘、气秘、风秘、热密、寒秘、大便不利、大便不通等字样者,或证候表现具有便秘症状者均在收录之列。对于方剂主治过于庞杂且组方用药对便秘缺乏针对性的条目,原则上不予以收录。

1.3 中药名称规范 参照《中药别名速查大辞典》^[5]对中药别名及俗称予以规范;参照全国高等中医药院校“十二五”规划教材《中药学》^[6]《中药大辞典》^[7]内容予以规范,并对其中中药功效进行归类。具体包括以下几个方面:1) 中药别名和俗称:如朴硝归为芒硝,锦纹归为大黄等;2) 中药合写或出现方剂名:如三白汤、三黄汤、栀子豉汤、六一散等将其补全;3) 中药炮制或产地名称:如茅苍术、北苍术归为苍术,汉防己归为防己,南木香归为木香,川柴胡归为

柴胡等;4) 药名书写不规范或方剂组成相同:如将诃子归为诃梨勒等。中药名称规范化处理见附表 1。

1.4 数据录入 将所纳入方剂按照表 1 格式进行录入,如果该方剂内含有此味中药则记录为“1”,反之,没有此味中药则为“0”。进行关联规则分析时,须将表 1 的数据格式进行转换,即将“1”“0”分别用“T”“F”替代并保存。

1.5 数据分析 应用频数及高频药物频率进行数据挖掘分析便秘方剂所有组成中药的构成情况,出现频数按照从高到低的顺序排序;应用因子分析挖掘高频药物的特征值、贡献率和综合得分;采用聚类分析中的系统聚类算法挖掘治疗便秘方剂的常用中药;采用关联规则分析中的 Apriori 算法挖掘高频中药药对之间的配伍关系。

表 1 数据录入格式举例

方剂	大黄	黄芩	葶苈子	防己	枳实	以后略
涤痰汤	1	1	0	0	0
润肠丸	0	0	1	1	1
消黄汤	1	1	0	0	0
以后略

2 结果

2.1 药物使用频次、频率分析 按照上述标准筛选后,得到明确可用于治疗便秘的方剂共 287 首,其中涉及中药 211 味,药物使用总频数达 2 535 次。通过频数频率统计,单味药中使用频率最高的是甘草,共使用 109 次,使用频率为 4.30%;大黄次之,共使用 103 次,使用频率为 4.06%。使用频率排列前 10 位的中药依次为:甘草、大黄、黄芩、当归、芍药、陈皮、木香、地黄、桂枝、枳实。以单味药使用频数 ≥ 平均频数(使用频数 ≥ 12 次)的 69 味中药为本次数据挖掘的高频药物。见表 2。

2.2 高频药物性味、归经分析 按照《中药学》与《中药大辞典》统计高频药物四气、五味及归经的频数频率。见表 4、表 5、表 6。其中,温性药物使用频率最高,达 26.64%;寒性药物次之,达 20.87%。五味属性中,辛味、苦味、甘味药物使用频数最高,均大于 500 次,其使用频率依次为 33.38%, 31.79%, 21.58%, 累计频率超过 80%。归经属性中排列前 5 位的为脾经、肺经、胃经、肝经、心经。

2.3 高频药类使用频数、频率分析 将高频药物按照《中药学》与《中药大辞典》的药物分类标准进行分类,共得到药物类别 16 项,总频数为 1 958 次。见表 3。由表 3 可见,补虚药使用频数最高,达 331 次,占总使用频数的 17.50%;解表药次之,使用频数达

305次,占总使用频数的16.13%;清热药使用频数达295次,占总使用频数的15.60%,位居第3位。使用频率排列前6的药类依次为:补虚药、解表药、清热药、理气药、化痰止咳平喘药、泻下药,在高频药物的累计频率中高达79.00%,可见历代治疗便秘的方剂涉及的药物种类、数量均较广泛,具体用药则较为集中。

表2 药物频数频率分布(使用频数≥12次)

药名	频数(次)	频率(%)	药名	频数(次)	频率(%)	药名	频数(次)	频率(%)
甘草	109	4.30	芒硝	31	1.22	桃仁	17	0.67
大黄	103	4.06	枳实	28	1.10	朱砂	17	0.67
黄芩	67	2.64	杏仁	28	1.10	香附	16	0.63
当归	65	2.56	麦冬	26	1.03	雄黄	16	0.63
芍药	53	2.09	麻黄	26	1.03	葛根	16	0.63
陈皮	52	2.05	白术	25	0.99	天麻	15	0.59
木香	48	1.89	青皮	22	0.87	乌头	14	0.55
地黄	43	1.70	荆芥	21	0.83	瓜蒌	14	0.55
桂枝	43	1.70	薄荷	21	0.83	玄参	14	0.55
枳壳	42	1.66	黄柏	20	0.79	郁金	13	0.51
人参	41	1.62	石膏	20	0.79	天南星	13	0.51
茯苓	40	1.58	生姜	20	0.79	知母	13	0.51
半夏	40	1.58	贝母	19	0.75	吴茱萸	13	0.51
栀子	38	1.50	巴豆	19	0.75	沉香	13	0.51
川芎	37	1.46	木通	18	0.71	三棱	12	0.47
柴胡	36	1.42	连翘	18	0.71	黄芪	12	0.47
厚朴	35	1.38	独活	18	0.71	滑石	12	0.47
羌活	35	1.38	细辛	18	0.71	天花粉	12	0.47
桔梗	34	1.34	升麻	18	0.71	苍术	12	0.47
槟榔	34	1.34	白芷	18	0.71	麝香	12	0.47
防风	33	1.30	干姜	17	0.67	白僵蚕	12	0.47
黄连	33	1.30	桑白皮	17	0.67	紫苏	12	0.47
附子	32	1.26	牡丹皮	17	0.67	前胡	12	0.47

注:治疗便秘方剂组成药物名称及使用频数详见附表2

表3 药类频数频率分布(使用频数≥12次)

序号	药类	中药	频数(次)	频率(%)
1	补虚药	甘草、黄芩、当归、芍药、人参、白术	331	17.50
2	解表药	桂枝、柴胡、羌活、防风	305	16.13
3	清热药	黄芩、地黄、栀子、黄连	295	15.60
4	理气药	陈皮、木香、枳实、香附	221	11.69
5	化痰止咳平喘药	半夏、桔梗、杏仁、青皮	189	9.99
6	泻下药	大黄、巴豆、芒硝、郁李仁	153	8.09
7	活血化瘀药	川芎、桃仁、郁金、三棱	79	4.18
8	利水渗湿药	茯苓、木通、茵陈、薏苡仁	70	3.70
9	温里药	附子、干姜、吴茱萸、丁香	63	3.33
10	化湿药	厚朴、苍术、藿香、砂仁	47	2.49
11	驱虫药	槟榔	34	1.80
12	祛风湿药	独活、乌头、秦艽、防己	32	1.69
13	平肝息风药	天麻、白僵蚕、羚羊角、全蝎	27	1.43
14	安神药	朱砂	17	0.90
15	杀虫止痒药	雄黄	16	0.85
16	开窍药	麝香、石菖蒲	12	0.63

表4 药物归经频数频率分布(使用频数≥100)

序号	归经	频数(次)	频率(%)
1	脾	1 035	17.83
2	肺	977	16.83
3	胃	819	14.11
4	肝	659	11.35
5	心	651	11.21
6	大肠	493	8.49
7	肾	365	6.29
8	胆	256	4.41
9	膀胱	205	3.53
10	心包	140	2.41
11	小肠	103	1.77
12	三焦	102	1.76

表5 药物四气频数频率分布(使用频数≥20)

序号	四气	频数(次)	频率(%)
1	温	563	26.64
2	寒	441	20.87
3	平	410	19.40
4	微寒	293	13.87
5	微温	253	11.97
6	热	63	2.98
7	凉	37	1.75
8	大热	33	1.56
9	大寒	20	0.95

表6 药物五味频数频率分布(使用频数≥30)

序号	五味	频数(次)	频率(%)
1	辛	1 050	33.38
2	苦	1 000	31.79
3	甘	679	21.58
4	酸	139	4.42
5	微苦	135	4.29
6	咸	57	1.81
7	淡	52	1.65
8	微甘	34	1.08

2.4 因子分析结果 运用SPSS 19.0统计软件对高频药物进行因子分析,首先进行KMO和Bartlett的检验。见表7。KMO统计量为0.715(>0.5),Bartlett球形检验 $P=0.000(< \approx 0.05)$,说明药物间相关程度较好,适合做因子分析。前25个公因子累计方差贡献率达71.141%,说明可以用该25个公因子进行代替。见表8。运用最大方差法(即具有Kaiser标准化的正交旋转法)进行旋转,旋转在25次迭代后收敛,得到相对较好效果,提取出各公因子中载荷量>0.4者,若提出的变量在多个公因子中同时出现,说明其不能作为公因子代表,宜排除,得出结果:黄连(0.694)、栀子(0.685)、黄柏(0.679)、天花粉(0.664)、连翘(0.565)、玄参(0.514)在第一

因子 P1 上有较高的载荷;白僵蚕(0.824)、天麻(0.784)、麝香(0.574)在第二因子 P2 上有较高的载荷;荆芥(0.729)、苍术(0.643)、葛根(0.632)在第三因子 P3 上有较高的载荷,以此类推,详见附表 3。由旋转成分矩阵可写出每个成分的因子表达式, $BX1 = 0.334 \times F1 + 0.111 \times F2 + 0.238 \times F3 - 0.173 \times F4 - 0.076 \times F5 + \dots + 0.180 \times F21 + 0.029 \times F22 + 0.182 \times F23 + 0.029 \times F24 - 0.211 \times F25$ 。以此类推,可以得到其余 BX2 ~ BX69 的表达式。依据个体因子 F1 中每个指标所对应的系数,得到个体因子 F1 的表达式, $F1 = -0.058 \times BX1 + 0.072 \times BF2 + 0.105 \times BX3 + 0.056 \times BX4 - 0.073 \times BX5 + \dots + 0.214 \times BX64 - 0.057 \times BX65 - 0.016 \times BX66 - 0.001 \times BX67 - 0.051 \times BX68 + 0.073 \times BX69$ 。以此类推,可以得到其余 F2 ~ F25 的表达式。碎石图见图 1,旋转载荷见图 2。

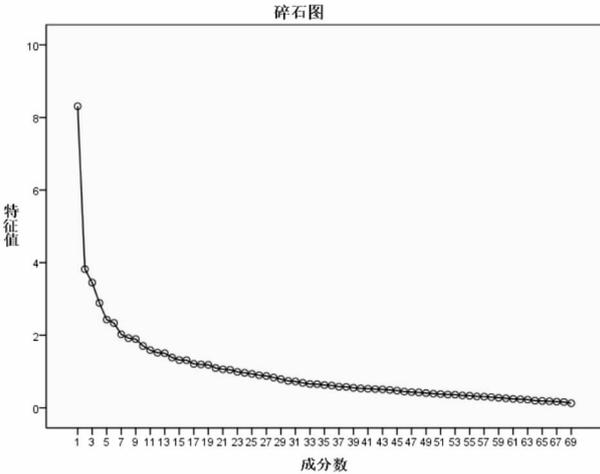


图 1 因子分析碎石图

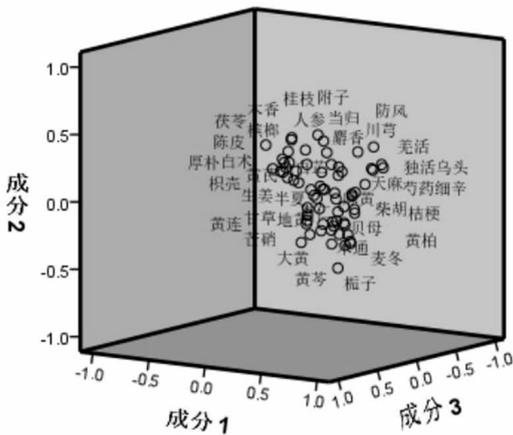


图 2 因子分析旋转载荷图

2.5 系统聚类分析结果 运用 SPSS 19.0 统计软件对高频药物进行系统聚类分析,聚类方法为 Ward 法,变量间距离定义使用平方 Euclidean 距离,聚类

结果见图 3。将结果聚为 5 大类较为合理,分别是第 1 类:大黄、甘草;第 2 类:1 组:当归、地黄、芍药;2 组:人参、白术、茯苓;3 组:羌活、防风、川芎、桂枝;第 3 类:1 组:黄连、黄柏;2 组:栀子,桔梗,柴胡;3 组:黄芩;第 4 类:1 组:陈皮、木香;2 组:香附、沉香、青皮、枳实、槟榔、厚朴;3 组:枳壳、半夏;第五类:1 组:芒硝、麦冬;2 组:其他。演化得到 10 个核心组合结果见表 9。

表 7 因子分析 KMO 统计量和球形检验

取样足够度的 Kaiser-Meyer-Olkin 度量。	0.715	
Bartlett 的球形度检验	近似卡方	7117.64
	df	2346
	Sig.	0

表 8 因子分析特征值及贡献率

成份	特征值	方差贡献(%)	累积方差贡献(%)
1	3.760	5.449	5.449
2	3.078	4.461	9.910
3	2.867	4.155	14.065
4	2.695	3.906	17.971
5	2.519	3.651	21.622
6	2.518	3.650	25.272
7	2.288	3.316	28.588
8	2.207	3.198	31.786
9	2.112	3.060	34.846
10	1.885	2.731	37.578
11	1.825	2.645	40.223
12	1.768	2.563	42.786
13	1.757	2.546	45.332
14	1.728	2.504	47.835
15	1.661	2.407	50.242
16	1.625	2.355	52.597
17	1.591	2.306	54.904
18	1.522	2.205	57.109
19	1.458	2.113	59.222
20	1.440	2.087	61.309
21	1.403	2.033	63.342
22	1.395	2.021	65.364
23	1.393	2.019	67.383
24	1.320	1.914	69.296
25	1.273	1.845	71.141

2.6 关联规则分析结果 运用 SPSS 14.2 统计软件进行高频药物的关联规则分析,设置参数:最低条件支持度 = 10%,最小规则置信度 = 55%,最终得到药物组合共 32 组,并据此分析结果进一步绘制出关联网络图。见表 10 和图 4。分析可知,在 2 味药的关联规则中,甘草→黄芩和大黄→黄芩的支持度均为 23.34%,并列第 1 位,其中甘草→黄芩的置信度更高,为 58.21%,大黄→黄芩的置信度相对稍低,为 56.72%;甘草→芍药的支持度为 18.47%,置信

度为 67.92%，排列第 3 位。在 3 味药的关联规则中，大黄→黄芩、甘草的支持度最高，达 18.47%，其置信度为 67.92%；栀子→黄芩、大黄的支持度次之，达 13.24%，其置信度为 65.79%。

表 9 治疗便秘方剂药物核心组合分析

序号	核心组合
1	大黄、甘草
2	当归、地黄、芍药
3	人参、白术、茯苓
4	羌活、防风、川芎、桂枝
5	黄连、黄柏
6	栀子、桔梗、柴胡
7	陈皮、木香
8	青皮、枳实、厚朴、沉香
9	枳实、半夏
10	细辛、白芷、生姜

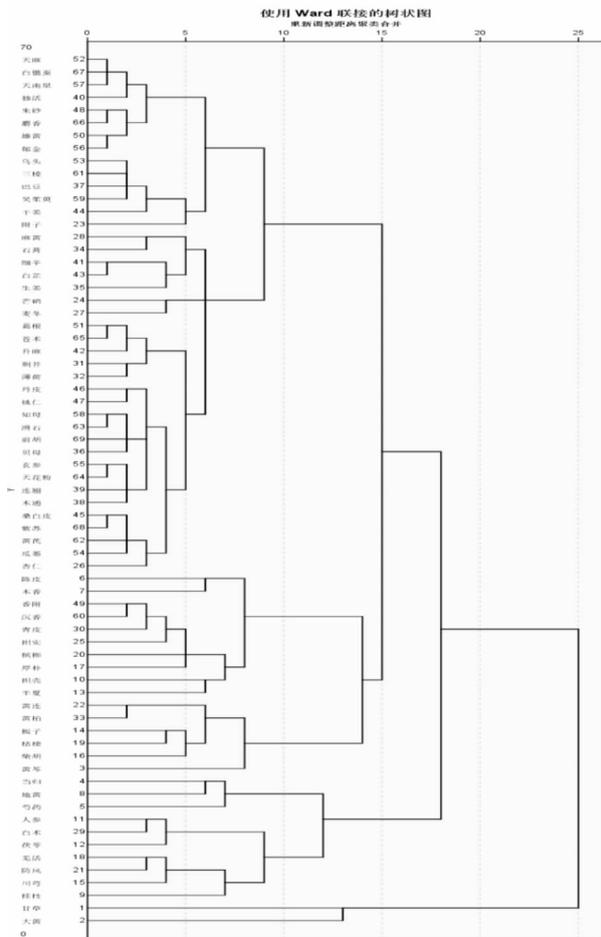


图 3 系统分析树状图

3 讨论

《兰室秘藏·大便结燥门》载：“若饥饱失节，劳役过度，损伤胃气，及食辛热味厚之物，而助火邪，伏于血中，耗散真阴，津液亏少，故大便结燥”；《扁鹊心书·便秘》载：“老人气虚，及妇人产后血少，致津

液不行，不得通流，故大便常结”。《名医杂著》载：“证属形气病，形气俱不足，脾胃虚弱，津血枯涸而大便难尔”。由此可见阴阳气血不足均是导致便秘的内在基础，感受外邪是便秘发生的外在条件，热结、气滞、寒凝、津液耗伤为其病因根本，病位首在大肠，并同时与肺、脾、胃、肝、肾等脏腑关系密切。本研究依据国家人口与健康科学数据共享平台提供的方剂专题服务数据，筛选出与便秘相关的 287 首方剂，采用频数分析、因子分析、系统聚类、关联规则分析等数据挖掘方法，运用 SPSS 19.0 统计软件与 SPSS 14.2 统计软件等数据挖掘软件总结出治疗便秘的方剂用药及配伍规律，以期为便秘的临床治疗提供思路。

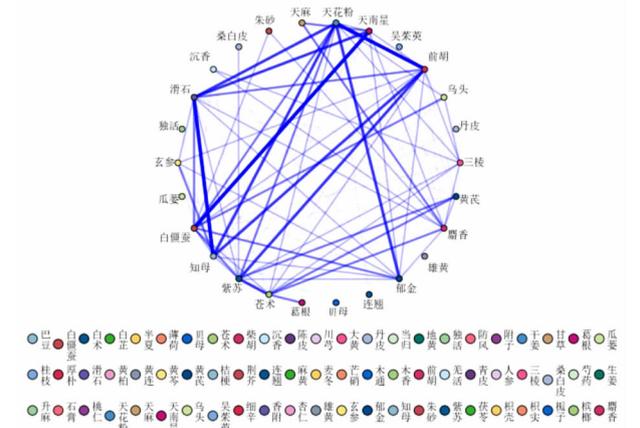


图 4 药物关联网络

表 10 药物关联规则分析

后项	前项	支持度 (%)	置信度 (%)	后项	前项	支持度 (%)	置信度 (%)
甘草	黄芩	23.34	58.21	甘草	柴胡	12.54	58.33
大黄	黄芩	23.34	56.72	当归	芍药、甘草	12.54	58.33
甘草	芍药	18.47	67.92	黄芩	芍药、甘草	12.54	55.56
甘草	陈皮	18.12	55.77	甘草	厚朴	12.20	57.14
当归	地黄	14.98	62.79	防风	羌活	12.20	57.14
甘草	桂枝	14.98	60.47	甘草	桔梗	11.85	67.65
甘草	茯苓	13.94	57.50	大黄	槟榔	11.85	61.76
大黄	黄芩、甘草	13.59	58.97	芍药	当归、甘草	11.85	61.76
黄芩	栀子	13.24	78.95	大黄	黄连	11.50	60.61
大黄	栀子	13.24	78.95	羌活	防风	11.50	60.61
栀子	黄芩、大黄	13.24	65.79	当归	防风	11.50	57.58
甘草	黄芩、大黄	13.24	60.53	大黄	芒硝	10.80	74.19
甘草	栀子	13.24	55.26	大黄	栀子、黄芩	10.45	83.33
甘草	川芎	12.89	62.16	黄芩	栀子、大黄	10.45	83.33
当归	川芎	12.89	56.76	甘草	栀子、黄芩	10.45	63.33
黄芩	柴胡	12.54	58.33	甘草	栀子、大黄	10.45	56.67

注：a. 支持度体现了药对或药组在方剂中是否为普遍存在的规律；如支持度为 30%，表示在 100 个方剂中有 30 个方剂符合关联规则描述；b. 置信度体现了关联规则前提成立的条件下药对或药组成果的概率；如置信度为 66%，表示在 100 个后项出现的方剂中有 66 个方剂包含前项

3.1 中药频数分析 通过对治疗便秘 287 首方剂所涉及到的 69 味高频药物用药分析可见,方剂中所涉及到的药物分类主要以补虚药、泻下药、理气药、解表药化痰平喘药以及活血化瘀药等为主。单味药使用频数 ≥ 50 的药物有甘草、大黄、黄芩、当归、芍药以及陈皮。上述药物分类以及单味药的使用基本能够体现历代医家对便秘相关用药规律的认识。饮食不节、劳役过度、脾胃气虚等导致大肠通降失调,造成传导不利是便秘的基本病机。在便秘的发展过程中,热盛津伤占有主要影响地位^[8]。此外,胃肠虚寒和肺脏功能失调亦是导致便秘的原因之一。因此,便秘治疗时以润肠清热为基本原则,又可根据病性的虚实,分别予以泻热、温散、通导、益气温阳或是滋阴养血之法。通过对 287 首治疗便秘方剂用药规律的研究显示,其结果恰如《景岳全书·秘结》曰:“阳结证,必因邪火有余,以致津液干燥。大肠热盛,燥屎内结。泻下热结,热去结散便通”。用药多以补阴药、泻下药为主。又如《伤寒名案选新注》载李士材医案:“治吴君明,伤寒六日,谵语狂笑,头痛有汗,大便不通,小便自利。方今仲冬,宜与桂枝汤……明日大便自通”巧用解表药治疗便秘^[9]。清·陈士铎《石室秘录·大便闭结》云:“大便闭结者,人以为大肠燥甚,谁知是肺气燥乎?肺燥则清肃之气不能下行于大肠”^[10]。明确指出便秘与肺有关,用药多以化痰平喘药与理气药等为主,契合便秘的治疗理念。就使用频数较高的药类分析来看,补虚药使用频次最高,其中以补气、补血药为主,代表药为当归、芍药、人参、白术、甘草等。解表药紧随其次,以桂枝、羌活和柴胡等药物为主。泻下药以大黄和芒硝等攻下药为主。清热药、理气药、化痰止咳平喘药、利水渗湿药以及活血化瘀药的使用频率也较高,具有一定的研究意义。因此便秘治疗总以扶正祛邪,恢复大肠传导功能为主。在便秘治疗中,补虚药与解表药占有重要地位,常配伍清热药,辅以理气、化痰、泻下药使用。

3.2 因子分析 因子分析的目标是通过发掘隐藏在数据下的一组较少的、更为基本的无法观测的变量,来解释一组可观测变量的相关性。这些虚拟的、无法观测的变量称作因子(每个因子被认为可解释多个观测变量间共有的方差,因此准确来说,他们应该称作公共因子)^[11]。利用因子分析的意义在于比较分析治疗便秘方剂中的药物组方规律,通过因子载荷系数筛选出在治疗便秘中发挥功效的主要中药类别及单味中药。本研究显示, KMO 统计量 =

0.715(满足 >0.5 的检测标准),由于提取主成分过多导致前 25 个主成分(F1 ~ F25)对应的特征根介于 1.273 与 3.760 之间,彼此区别不是很大,不能很好的解释原始变量,概其原由是治疗便秘方剂过于庞杂所致。参照附表 3,对 25 个主成分初步提取,例如黄连(0.694)、栀子(0.685)、黄柏(0.679)、天花粉(0.664)、连翘(0.565)、玄参(0.514)在 F1 上有较高的载荷;白僵蚕(0.824)、天麻(0.784)、麝香(0.574)在 F2 上有较高的载荷;荆芥(0.729)、苍术(0.643)、葛根(0.632)在 F3 上有较高的载荷;陈皮(0.758)、青皮(0.718)在 F5 上有较高载荷。从药类使用上分析,基本契合治疗便秘的常用药类,即以解表药、祛风湿药、补虚药、理气药、活血化瘀药、清热药、平肝熄风药、化痰止咳平喘药以及温里药等为主。因此,作为一种数学计算方法,因子分析在治疗便秘方剂的客观的定量评价方面具有优势,从另一个侧面为中医药治疗便秘提供了一个新的尝试性研究方法,为便秘方剂的临床应用提供了参考性指标。

3.3 系统聚类分析 系统聚类是实际工作中使用的最多的一种聚类方法,它具有十分明显的优点:可以对样品聚类,也可以对变量聚类。通过将 n 个样品或变量看成不同的 n 类,然后将距离接近(针对样品聚类)或性质接近(针对变量聚类)的两类合并为一类;再从这 $n-1$ 类中找到最接近的 2 个类合并,以此类推,直到所有的样品或变量被合为一类。整个过程可以绘成聚类图,按图和具体问题来决定分类^[12]。本研究运用系统聚类分析法,在通过对治疗便秘方剂数据的多次聚类后,结合治疗便秘方剂的药物组成分析,系统聚类模型显示,治疗便秘方剂比较好的聚类数目为 5 类。分析可见,第 1 类方剂主要由大黄、甘草等药物组成,聚一类以泻下药与补虚药相配伍,主要针对热结下焦与脾胃气虚两类病理因素;第 2 类方剂组成有当归、地黄、芍药均为补血药,针对阴血虚,润泽荣养不足所致便秘;人参、白术、茯苓为补气药与利水渗湿药的组合,针对气虚水停,温煦传送无力所致便秘;羌活、防风、桂枝、川芎为发散风寒药与活血化瘀药组合,主要针对外感寒邪与瘀血停滞两类病理因素;第 3 类方剂主要组成有黄连与黄柏清热燥湿药配伍,针对湿热雍滞,气机内停所致便秘;栀子,桔梗,柴胡为清热泻火药、清化热痰药与发散风热药组合,针对热盛津伤所致便秘;第 4 类方剂组成为陈皮、木香为理气药组合,针对情志失调,气机郁滞的病理因素;香附、沉香、青皮、枳实、厚朴、槟榔为理气药与化湿药组合,针对湿气瘀

滞,气机不通所致便秘;枳实、半夏为理气药与温化寒痰药组合,针对寒凝气滞,痰饮内停两类病理因素;第5类中药聚类相效果不理想,但其中芒硝与麦冬为攻下药与补阴药的组合,针对津液亏损,大肠失润所致便秘。结果显示系统聚类聚为5类时,5类治疗便秘方剂中主要药物组成虽有很多相似之处,但第1、2类方剂的药物组成以常用药物为主,而第3、4类方剂的药物组成中除了常用药物还有许多特殊药物。经与相关中医传统理论及其他相似研究对比,可以认为由系统聚类算法所确定的5类药物组成与已有的理论研究及临床研究结果之间符合度较好,不存在明显矛盾,并且在进行数据挖掘时还总结出治疗便秘的特殊药物有陈皮、香附、桔梗、半夏、麝香、郁金、乌头、三棱、吴茱萸等,对治疗便秘方剂的理论与实践研究具有一定的推动作用。其次,基于聚类分析,本研究演化得到治疗便秘的核心组合10个,分别为大黄、甘草;当归、地黄、芍药;人参、白术、茯苓;羌活、防风、川芎、桂枝;黄连、黄柏;栀子、桔梗、柴胡;陈皮、木香;青皮、枳实、厚朴、沉香;枳实、半夏;细辛、白芷、生姜,这10个核心组合同样印证饮食不节、情志失调、年老体虚、感受外邪会造成热结、气滞、寒凝、气血阴阳亏虚,对便秘的发生发展不容忽视。

3.4 关联规则分析 关联规则又称关联挖掘,能够反映两药之间的配伍紧密程度、相关性或因果结构,依赖关系网络分析可以发现某些属性同时出现的规律和模式,进而发现核心药物,两者结合则可以发现核心药组^[13]。本研究为探索治疗便秘运用Apriori算法对高频药物进行关联规则分析,为探索治疗便秘方剂中新的组方规律,本研究将设置参数调为支持度 $\geq 10\%$,置信度 $\geq 55\%$,共得出32条二元矩阵的药对与药组关联规则。见表10。结果分析可见,大黄→黄芩、栀子;大黄→芒硝;甘草→栀子、黄芩;大黄→黄连;黄芩→栀子;大黄→甘草;黄芩→甘草为已知治疗便秘常用药对^[14]。由此,我们更进一步将设置参数调为支持度 $\geq 10\%$,置信度 $\geq 60\%$,共得出17条药对与药组关联规则,支持度较高的药物组合有大黄→栀子;甘草→黄芩;甘草→大黄;甘草→川芎。这些药对与药组主要是补虚药、活血化瘀药、清热药、泻下药的不同组合,如泻下药→清热药(大黄→栀子);补虚药→清热药(甘草→黄芩);补虚药→活血化瘀药(甘草→川芎);补虚药→泻下药(甘草→大黄)。上述结果体现了治疗便秘的泻热、温散、通导、益气温阳与滋阴养血等治法。

大黄→槟榔;当归→地黄;甘草→陈皮;羌活→防风;芍药→当归是现今尚未总结的隐性结论,值得进行进一步讨论。

3.5 创新点与研究不足

3.5.1 创新点分析 由治疗便秘组方为切入点,对《中国方剂数据库》相关内容进行了系统梳理并建立治疗便秘的2次数据库,综合运用Excel 2016、SPSS 19.0统计软件与SPSS 14.2统计软件等数据挖掘软件,运用因子分析、频数统计、系统聚类分析及关联规则分析等手段,依据药物类别→单味药→药对或药组→方剂的顺序,从局部到整体的不同角度进行研究治疗便秘组方的药类分析和用药规律,探讨其功效、主治、药对配伍等相关理论,验证部分中药在治疗便秘方剂配伍研究中的合理性与可行性,为发挥中医药治疗优势,提高便秘的临床治疗效果提供可靠的数据支持。此外,通过聚类分析结果显示,治疗便秘的核心类别有10类,分别为大黄、甘草;当归、地黄、芍药;人参、白术、茯苓;羌活、防风、川芎、桂枝;黄连、黄柏;栀子、桔梗、柴胡;陈皮、木香;青皮、枳实、厚朴、沉香;枳实、半夏;细辛、白芷、生姜,同时也印证了补虚药、解表药、泻下药以及理气药对于便秘治疗的有效性。

3.5.2 研究不足分析 本研究的数据受国家人口与健康科学数据共享平台提供的《中国方剂数据库》支持,其中共收录了万余首方剂数据,符合数据挖掘技术对数据源数据量大的基本要求。在研究过程中,我们发现了存在两首方剂的名称相同的情况。通过查阅古籍发现,方剂数据库录入的主治与古籍记载不完全一致等情况,因此可能会对数据挖掘研究造成偏倚,从而影响结论的可靠性。其次,从数据挖掘角度考虑,如果能收集到更多治疗便秘的方剂信息并做好规范化数据处理,对治疗便秘方剂的数据挖掘就能更加全面、更加准确。此次对治疗便秘方剂的数据挖掘主要选择了频数分析、因子分析、系统聚类分析、关联规则分析等数据挖掘方法,对于发现频率较高的隐含数据信息效果较好,但无法研究出现频率较低的信息。由此可见,临床实践上由医家对某方面的独特认识而产生的一些特殊的数据项,尽管其出现频率不高,但确实重要,不能被忽视。基于关联规则分析,我们虽然也试图通过降低最小支持度来减小算法本身的缺陷,但不可否认的是这依然是目前所存在的一大难题。

4 小结

本研究通过对方剂专题服务数据中治疗便秘方

剂进行筛选,得到二次数据库,采用频数分析、因子分析、系统聚类分析和关联规则分析等数据挖掘方法,对治疗便秘的方剂进行数据挖掘分析。

结果显示,补虚药、泻下药、理气药、解表药化痰平喘药以及活血化瘀药在便秘的使用中占有较大比重。便秘治疗总以扶正祛邪,恢复大肠传导功能为主。治疗时以润肠清热为基本原则,又可根据病性的虚实,分别予以泻热、温散、通导、益气温阳或是滋阴养血之法。进行数据挖掘时还总结出治疗便秘的特殊药物有陈皮、香附、桔梗、半夏、麝香、郁金、乌头、三棱、吴茱萸等。大黄→槟榔;当归→地黄;甘草→陈皮;羌活→防风;芍药→当归是现今尚未总结的隐性结论,值得进行进一步讨论。数据挖掘适用于治疗便秘方剂的研究,我们通过数据挖掘技术,量化地挖掘出便秘常用药物、特殊药物以及药对,以期通过直观数据的形式为治疗便秘提供临床参考,也可为科研提供研究思路。

5 致谢

本组全体成员对提供此次比赛机会的主办方国家科技基础条件平台中心表示由衷的感谢,同时,感谢国家人口与健康科学数据共享平台和中国中医科学院中医药信息研究所提供的中国方剂数据库,感谢 CNKI 提供的文献查询支持,感谢本组此次比赛中所引用书籍、文献的主编及作者,感谢各位评委老师对本组作品的悉心评阅,感谢学校给予浓厚的学术氛围。最后,再一次真诚的感谢我们的父母以及关心、帮助过我们的老师和同学,他们宝贵的意见是我们灵感和动力的源泉。

参考文献

[1] 吴勉华,王新月. 中医内科学[M]. 北京:中国中医药出版社, 2012.

[2] 李佳丽,颜立志,季田,等. 主成分分析古医籍中肝癌类证候组方规律[J]. 中国中西医结合外科杂志,2016,22(6):523-528.

[3] 许仲宇. 系统聚类关联分析大肠癌放射性肠炎组方规律[J]. 光明中医,2018,33(5):753-754,后插1,封3.

[4] 满达,吴超群,艳红,等. 基于关联分析的蒙医萨病方剂组方规律研究[J]. 中国民族医药杂志,2017,23(7):46-47.

[5] 李顺保. 中药别名速查大辞典[M]. 北京:学苑出版社,1997:65.

[6] 陈蔚文. 中药学[M]. 2版. 北京:人民卫生出版社,2013:73.

[7] 江苏新医学院. 中药大辞典[M]. 上海:上海人民出版社,1977:87.

[8] 李霞,朱思敏,李防栋. 刘绍能主任治疗便秘经验[J]. 医学信息, 2015,28(41):362-362.

[9] 张磊. 舒芬太尼麻醉用于心血管手术的多中心临床研究[J]. 中国卫生产业,2014,12(3):108-109.

[10] 阳婷婷,唐铁洋,张兴安. 舒芬太尼与瑞芬太尼复合丙泊酚全凭静脉麻醉用于神经外科手术效果与安全性的系统评价[J]. 中国循证医学杂志,2013,13(12):1492-1499.

[11] 吴化宇,丁婕,雒朝君,等. 基于验证性因子分析的中医工作满意度量表研制[J]. 中国卫生统计,2019,36(1):92-95.

[12] 潘发明. SPSS 统计分析大全[M]. 北京:清华大学出版社,2014:328-329.

[13] 陈文莉,李文林,史勤,等. 基于关联规则及复杂网络分析古医籍中乳腺癌的用药规律[J]. 南京中医药大学学报,2018,34(6):625-629.

[14] 石强. 基于数据挖掘对柴胡类方证规律的研究[D]. 北京:北京中医药大学,2012.

(2018-10-30 收稿 责任编辑:杨觉雄)

附表1 中药名称规范化处理

中药	中药别名及俗称	中药	中药别名及俗称
艾叶	艾绒	甘草	粉草
白及	白芨	干姜	白姜
白藜	白敛	葛根	干葛
白术	冬术	荆芥	荆穗
蚕砂	蚕沙	铅丹	黄丹
蚕蜕	蚕退	生姜	黄姜
苍术	仙术	续断	川断
蝉蜕	蝉退	玉竹	女萎
蟾酥	干蟾	珍珠	真珠
陈皮	广皮	栀子	山栀
大黄	熟军	麦芽	生谷芽
槟榔	白槟榔	木瓜	宣木瓜
当归	西当归	木香	南木香
诃子	诃子皮	天冬	天门冬
黄芩	淡黄芩	茺萸	白茺萸
僵蚕	白僵蚕	芦根	水芦根笋
连翘	连翘壳	鼠妇	螂(虫祁)
冰片	片脑、龙脑	神曲	法曲、麦曲
川芎	芎藭、芎藭	朱砂	丹砂、辰砂
花椒	川椒、蜀椒	茯苓	白茯苓、云苓
天麻	赤箭、明天麻	芥子	白芥子、芥菜子
白芷	吴白芷、杭白芷	血余	男人血余、发团
茴香	大茴香、小茴香	大枣	广枣、小红枣、枣膏
全蝎	干蝎、蝎梢、蝎尾	附子	天雄、黑附子、生附子
乌头	侧子、川乌、草乌	密陀僧	陀僧
白茅根	茅根	牡丹皮	丹皮
白石英	石英	牛蒡子	粘子
侧柏叶	柏叶	青娘子	芫青
海蛤壳	海蛤	乌梢蛇	乌蛇
龙胆草	龙胆	莱菔子	萝卜子
巴戟天	紫巴戟	山慈菇	毛慈姑
赤芍药	红芍药	葶苈子	苦葶苈
核桃仁	胡桃肉	吴茱萸	食茱萸
葫芦巴	葫芦巴	淫羊藿	仙灵脾
闹羊花	踯躅花、踯躅	天南星	南星、生南星
牵牛子	黑丑头、白牵牛	大风子	大枫子、大枫子肉
延胡索	延胡、胡锁、玄胡索	猪牙皂	牙皂、皂荚、皂角、猪牙皂角

附表2 药物频数频率分布总表

中药名称	频数(次)	频率(%)	中药名称	频数(次)	频率(%)	中药名称	频数(次)	频率(%)
甘草	109	4.299	地骨皮	11	0.433	芫花	4	0.157
大黄	103	4.063	金银花	10	0.394	扁豆	4	0.157
黄芩	67	2.642	藿香	10	0.394	乌梢蛇	4	0.157
当归	65	2.564	郁李仁	10	0.394	姜黄	4	0.157
芍药	53	2.090	砂仁	10	0.394	苦参	4	0.157
陈皮	52	2.051	秦艽	10	0.394	射干	3	0.118
木香	48	1.893	羚羊角	9	0.355	柏子仁	3	0.118
地黄	43	1.696	防己	9	0.355	木鳖子	3	0.118
桂枝	43	1.696	神曲	9	0.355	泽兰	3	0.118
枳壳	42	1.656	丹参	9	0.355	芦荟	3	0.118
人参	41	1.617	麻子仁	9	0.355	百合	3	0.118
茯苓	40	1.577	全蝎	9	0.355	大青叶	3	0.118
半夏	40	1.577	茴香	9	0.355	紫草	3	0.118
栀子	38	1.499	石菖蒲	9	0.355	川椒	3	0.118
川芎	37	1.459	丁香	9	0.355	五灵脂	3	0.118
柴胡	36	1.420	牛黄	9	0.355	琥珀	3	0.118
厚朴	35	1.380	紫菀	9	0.355	蒲公英	3	0.118
羌活	35	1.380	天门冬	8	0.315	硫磺	3	0.118
桔梗	34	1.341	牵牛子	8	0.315	山药	3	0.118
槟榔	34	1.341	茵陈	8	0.315	木瓜	3	0.118
防风	33	1.301	薏苡仁	8	0.315	香薷	3	0.118
黄连	33	1.301	甘遂	8	0.315	寒水石	3	0.118
附子	32	1.262	五味子	8	0.315	益母草	3	0.118
芒硝	31	1.222	皂角	8	0.315	山豆根	3	0.118
枳实	28	1.104	大枣	8	0.315	续断	3	0.118
杏仁	28	1.104	龙脑	7	0.276	花椒	3	0.118
麦冬	26	1.025	苏叶	7	0.276	忍冬藤	3	0.118
麻黄	26	1.025	牛膝	7	0.276	钩藤	3	0.118
白术	25	0.986	菊花	7	0.276	土瓜根	3	0.118
青皮	22	0.867	草果	7	0.276	旋覆花	3	0.118
荆芥	21	0.828	青蒿	7	0.276	款冬花	3	0.118
薄荷	21	0.828	荷叶	7	0.276	干漆	2	0.078
黄柏	20	0.788	车前子	7	0.276	巴戟天	2	0.078
石膏	20	0.788	石斛	6	0.236	韭白	2	0.078
生姜	20	0.788	麦芽	6	0.236	白部	2	0.078
贝母	19	0.749	山楂	6	0.236	益智仁	2	0.078
巴豆	19	0.749	鳖甲	6	0.236	桑寄生	2	0.078
木通	18	0.710	红花	6	0.236	冬瓜仁	2	0.078
连翘	18	0.710	豆豉	6	0.236	白藜皮	2	0.078
独活	18	0.710	橘红	6	0.236	杜仲	2	0.078
细辛	18	0.710	豆蔻	6	0.236	苦楝皮	2	0.078
升麻	18	0.710	龙胆草	6	0.236	蛇蜕	2	0.078
白芷	18	0.710	泽泻	6	0.236	赤小豆	2	0.078
干姜	17	0.670	乌梅	5	0.197	竹茹	2	0.078
桑白皮	17	0.670	葶苈子	5	0.197	天竺黄	2	0.078
牡丹皮	17	0.670	使君子	5	0.197	薤白	2	0.078
桃仁	17	0.670	决明子	5	0.197	阿胶	2	0.078
朱砂	17	0.670	胡椒	5	0.197	桑叶	2	0.078
香附	16	0.631	牛蒡子	5	0.197	芫蕘	2	0.078
雄黄	16	0.631	河黎勒皮	5	0.197	白前	2	0.078
葛根	16	0.631	高良姜	5	0.197	篇蓄	1	0.039
天麻	15	0.591	乳香	5	0.197	瞿麦	1	0.039

续附表2 药物频数频率分布总表

中药名称	频数(次)	频率(%)	中药名称	频数(次)	频率(%)	中药名称	频数(次)	频率(%)
乌头	14	0.552	没药	5	0.197	五加皮	1	0.039
瓜蒌	14	0.552	粳米	5	0.197	海金沙	1	0.039
玄参	14	0.552	蝉蜕	5	0.197	败酱草	1	0.039
郁金	13	0.512	威灵仙	5	0.197	白花蛇	1	0.039
天南星	13	0.512	大戟	5	0.197	蛇草	1	0.039
知母	13	0.512	葱白	5	0.197	党参	1	0.039
吴茱萸	13	0.512	穿山甲	5	0.197	玉竹	1	0.039
沉香	13	0.512	礞砂	4	0.157	龟甲	1	0.039
三棱	12	0.473	白鲜皮	4	0.157	水蛭	1	0.039
黄芪	12	0.473	大腹皮	4	0.157	虻虫	1	0.039
滑石	12	0.473	延胡索	4	0.157	灵芝	1	0.039
天花粉	12	0.473	淡竹叶	4	0.157	核桃仁	1	0.039
苍术	12	0.473	川楝子	4	0.157	狗脊	1	0.039
麝香	12	0.473	蒲黄	4	0.157	蜈蚣	1	0.039
白僵蚕	12	0.473	代赭石	4	0.157	蛤蚧	1	0.039
紫苏	12	0.473	青黛	4	0.157	茶叶	1	0.039
前胡	12	0.473	蒺藜	4	0.157	艾蒿	1	0.039
莪术	11	0.433	远志	4	0.157	沙参	1	0.039
犀角	11	0.433	白矾	4	0.157			

附表3 正交旋转因子分析表

公因子	成分
F1	黄连(0.694)、栀子(0.685)、黄柏(0.679)、天花粉(0.664)、连翘(0.565)、玄参(0.514)
F2	白僵蚕(0.824)、天麻(0.784)、麝香(0.574)
F3	荆芥(0.729)、苍术(0.643)、葛根(0.632)
F4	白芷(0.840)、细辛(0.672)、川芎(0.510)、羌活(0.480)、防风(0.409)
F5	陈皮(0.758)、青皮(0.718)
F6	桑白皮(0.688)、紫苏(0.676)、杏仁(0.625)、枳壳(0.576)
F7	当归(0.726)、地黄(0.723)、芍药(0.524)、黄芪(0.495)
F8	知母(0.782)、贝母(0.688)、牡丹皮(0.423)
F9	人参(0.801)、白术(0.664)、茯苓(0.648)
F10	柴胡(0.771)、黄芩(0.493)
F11	麻黄(0.751)、独活(0.441)
F12	麦冬(0.721)、芒硝(0.613)
F13	滑石(0.713)、木通(0.537)、天南星(0.436)
F14	吴茱萸(0.796)、厚朴(0.624)
F15	附子(0.654)、桂枝(0.543)、乌头(0.533)、干姜(0.477)
F16	雄黄(0.817)、朱砂(0.537)、郁金(0.420)
F17	枳实(0.792)
F18	巴豆(0.761)、乌头(0.463)、三棱(0.443)
F19	沉香(0.748)、木香(0.552)
F20	生姜(0.818)、半夏(0.531)
F21	槟榔(0.684)
F22	甘草(0.753)
F23	桃仁(0.865)
F24	石膏(0.758)
F25	瓜蒌(0.769)