

当归芍药散组方药味关联规律的数据挖掘研究

许 钊^{1,2} 彭代银² 余伯阳^{1#}

(1 中国药科大学中药复方研究室 江苏 210038; 2 安徽中医学院药学院)

摘要: **目的** 研究当归芍药散组方药味关联规律,为进一步研究当归芍药散组方原理提供基础。**方法** 通过检索中药方剂数据库中与当归芍药散组方药味有关的方剂,构建当归芍药散药味配伍数据库,运用神经网络、决策树、关联规则、聚类分析等数据挖掘技术,利用 SPSS Clementine 和 SPSS 软件,研究其组方药味之间的关联规律。**结果** 相互支持度,当归对芍药、芍药对当归为强支持,茯苓对白术、川芎对当归、当归对川芎、白术对茯苓、泽泻对茯苓为较强支持;对全方的贡献度,当归 > 茯苓 > 芍药 > 川芎 > 白术 > 泽泻;聚类分析表明泽泻对当归芍药散全方关联最强。**结论** 在当归芍药散组成中,当归、茯苓和芍药为主要成分,泽泻为特征成分,川芎、白术在其中发挥重要作用。**关键词:** 当归芍药散;组方;药味;关联规律;数据挖掘
中图分类号: R2-03

Data mining in relative rules among ingredients of Dangguishaoyao Powder

XU Fan^{1,2}, PENG Daiyin², YU Boyang^{1#}

(1 Research Room of Chinese Compound Formulas China Pharmaceutical University, Jiangsu 210038;
2 School of Pharmacy Anhui College of Chinese Medicine)

Abstract: **Objective** To study the relative rules among the ingredients of Dangguishaoyao Powder and provide a basis for further research. **Method** The relative rules among the ingredients of Dangguishaoyao Powder were studied by applying data mining technology including neural networks decision tree association rules and clustering analysis and software of SPSS Clementine and SPSS. The database of ingredient combinations of Dangguishaoyao Powder was established through retrieving the formulas related to the ingredients of Dangguishaoyao Powder from the Database of Chinese Formulas. **Result** In the aspect of mutual support rating Radix Angelicae Sinensis to Radix Paeoniae Alba and Radix Paeoniae Alba to Radix Angelicae Sinensis showed a stronger support rating. Poria to Rhizoma Atractylodis Alba Rhizoma Ligutici to Radix Angelicae Sinensis Radix Angelicae Sinensis to Rhizma Ligutici Rhizma Atractylodis Alba to Poria and Rhizma Alismatis to Poria showed a strong support rating. In the aspect of contribution rating to the whole formula Radix Angelicae Sinensis > Poria > Radix Paeoniae Alba > Rhizma Ligutici > Rhizma Atractylodis Alba > Rhizma Alismatis. The outcomes of clustering analysis showed that Rhizma Alismatis had the strongest relevancy to the whole formula of Dangguishaoyao Powder. **Conclusion** In Dangguishaoyao Powder Radix Angelicae Sinensis Poria and Radix Paeoniae Alba are the main ingredients Rhizma Alismatis is characterized ingredient and Rhizma Ligutici and Rhizma Atractylodis Alba play important roles.

Key words: Dangguishaoyao Powder; formula forming; formula ingredients; association rules; data mining

当归芍药散出自张仲景所著《金匱要略》,由当归、芍药、川芎、茯苓、泽泻、白术组成,用于治疗“妇

人怀妊,腹中疝痛”及“妇人腹中诸疾病”,被广泛用于临床各科,治疗效果显著。组成当归芍药散的 6

许 钊,男,在读博士生,副教授

#通讯作者:余伯阳,博士,教授,博士生导师, E-mail boyangu59@163.com

味药物,在浩瀚的中药和方剂的数据海洋中,可以以各种组合形式存在,或单用,或合用,或独立成方,或为其他方剂的组成部分。6味药物经仲景之手组成了当归芍药散,这显然不是一个随机事件,而有其内在规律和逻辑联系。数据挖掘是多学科技术的集成^[1]。借助数据挖掘技术对当归芍药散组方原理进行初步研究,将有利于深入了解该方的内涵,并为进一步的实验研究提供基础。

1 资料来源

药品资讯网(原新药研发咨询中心)建立的中药方剂数据库(www.chendrug.com/Prescription.asp)收录古今名方、验方、偏方等共计 10 704 个。

2 方法与结果

2.1 数据库建立

利用中药方剂数据库,将当归芍药散组方药味当归、芍药、川芎、白术、茯苓、泽泻分别输入进行检索并统计。应用 Microsoft Office Access 2003 建立当归芍药散药味配伍数据库,字段名包括方名、来源、药味名、药味配伍类型、主治等。

2.2 数据统计与数据挖掘研究

对数据库中当归芍药散各药味之间相互配伍的各种形式($2^6 - 1 = 63$ 种)出现的概率信息进行整理,并用 SPSS 软件进行因子分析和聚类分析;同时应用 Clementine 数据挖掘系统,采用神经网络和广义规则归纳法(GR 算法),计算当归、芍药、川芎、茯苓、泽泻、白术之间的相互关联(支持度和置信度)。

2.3 数据整理

从数据库中共检出含有当归芍药散药味的方剂 7 293 个,除去重复方剂,共有有效数据 3 942 条。当归、芍药、白术、泽泻与其他药味配伍,芍药、川芎、白术、泽泻和其他药味配伍,芍药、川芎、白术、茯苓、泽泻和其他药味配伍,3 种类型没有,故在全部 63 种配伍类型中,实际配伍类型共计 60 种。将方剂库中所有符合上述配伍类型的方剂检出并统计,考察各种配伍类型出现的几率和药味组成情况。

2.4 数据挖掘研究

2.4.1 数据预处理:将数据格式化。如当归白术汤,原方组成为:当归、白术、茯苓……,不含芍药、川芎、泽泻,则按当归、芍药、川芎、白术、茯苓、泽泻的顺序,将当归白术汤记为: {1, 0, 0, 1, 1, 0}。同理,将所有检索出的数据格式化,便于进一步挖掘。

2.4.2 神经网络和 GR 算法计算当归芍药散药味之间的相互关联结果:应用 Clementine 神经网络算

法和 GR 计算 6 味药彼此的相互支持度。运算流程如下(以当归为例):①分析:估计的准确性, 75.501; 输入层, 5 个神经元;隐藏层 1.20 个神经元;隐藏层 2.15 个神经元;隐藏层 3.10 个神经元;输出层, 1 个神经元。②字段:目标,当归;输入,芍药、川芎、白术、茯苓、泽泻。③构建设置:方法,快速;样本(%), 100.0;算法,神经网络、GR;模型类型,分类。④输入的相对重要性:芍药, 0.415 937;川芎, 0.353 021;茯苓, 0.283 897;白术, 0.238 492;泽泻, 0.201 752。同上,我们分别求出 6 味药彼此的相互支持度,并以支持度高低计分,估算出每味药对其他 5 味药的总支持度,初步判断该味药对全方的贡献度。①相互支持度:强支持($\geq 40\%$),当归对芍药、芍药对当归;较强支持($\geq 30\%$),茯苓对白术、川芎对当归、当归对川芎、白术对茯苓、泽泻对茯苓。②对全方的贡献度,当归 > 茯苓 > 芍药 > 川芎 > 白术 > 泽泻。

2.4.3 因子分析和聚类分析:假设所有与当归芍药散组成药味有关的方剂为 1 个集合 {当归芍药散}。文中集合共有 3 942 个事务,并分别归属于 60 个各自独立的子集 {当归}、……、{当归,芍药,川芎}、…、{当归,芍药,川芎,白术,茯苓,泽泻}。找出 {当归芍药散} 的共性规律,进而研究由这些规律赋予子集的属性(类别)。将 {当归芍药散} 中 3 942 个事务全面纳入运算,并通过 SPSS 进行因子分析、聚类分析。①因子分析:采用主成分分析法、提取特征值大于 1 的因子、进行方差最大正交旋转,将因子得分作为新变量存入。共提取出 6 个因子,到第 5 个因子可以解释 88.861% 的事务,为了全面反映所有事务的特点,我们提取全部 6 个因子。引入上述 6 个因子对数据进行校正,如补肝散原数据格式为: {1, 0, 1, 1, 0, 0}, 校正后为: {0.988 25, -1.024 74, 1.765 46, 1.747 39, -0.779 72, -0.175 02}。校正后的数据从离散形式(0 或 1),转换为连续数据,进行聚类分析。②聚类分析:引入聚类分析的 6 个变量(即上面的 6 个因子),对所有事务进行聚类分析,共聚为 6 类:第 1 类,主要解释当归、茯苓,部分解释芍药,排除川芎、泽泻;第 2 类,主要解释白术,部分解释芍药,排除川芎、泽泻;第 3 类,主要解释泽泻,不排除其他药味;第 4 类,主要解释川芎,部分解释芍药、当归,排除泽泻;第 5 类,主要解释泽泻,与第 3 类的区别在于白术增加,当归、茯苓减少;第 6 类,主要解释川芎,排除芍药、泽泻。结果见表 1。子集 {当归,芍药,川芎,白术,茯苓,泽泻},即当归芍药散全方属于第 3 类。

表 1 各类别方剂药味组成情况情况(个)						
类别	当归	芍药	川芎	白术	茯苓	泽泻
1	1 062	467	0	107	830	0
2	134	495	0	687	291	0
3	79	38	17	14	137	167
4	425	470	510	148	165	0
5	8	24	7	91	67	162
6	158	0	420	56	56	0

2.4.4 建模:已知当归芍药散全方属于第 3 类,我们重点讨论第 3 类的生成规则。应用 SPSS Clementine 运用关联规则、决策树等算法,训练数据库,生成规则(见图 1)。从图 1 可看出进入第 3 类则必须同时满足下列 2 个条件:①必须有泽泻;②当归和茯苓同时存在,或当归和茯苓只有 1 个但没有白术。上述条件是构成当归芍药散的必要条件,即满足上述条件不一定就是当归芍药散,但不满足肯定不是当归芍药散。

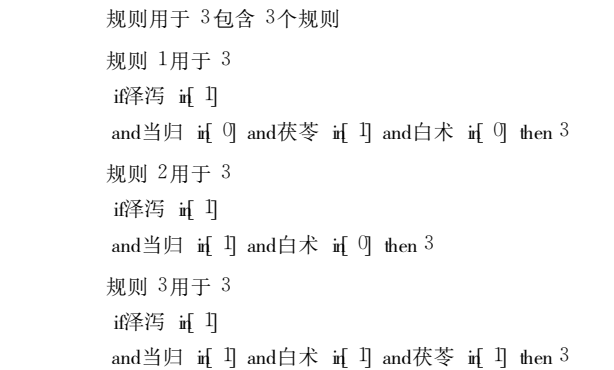


图 1 SPSS Clementine 生成的指向第 3 类的规则

3 讨论

从相互支持度结果可以发现,支持度高的当归与芍药、茯苓与白术、当归和川芎等均是临床中药方剂常用的药对或配伍形式,可见相互支持度结果是可信的。6 味药对全方的贡献度结果表明每味药对全方的贡献度,这个结果与表 1 的顺序基本一致,即药味出现的频率越高,其贡献度越高。从聚类分析结果可以看出,要满足当归芍药散同时具备 6 个变量的条件,则只有第 3 类和第 5 类。而第 5 类由于白术增加,当归、茯苓减少,同时泽泻的权重达 45%,与当归芍药散原方发生背离,因此当归芍药散属于第 3 类是可信的,同时泽泻对当归芍药散全方关联最强。综上所述,在当归芍药散组成成分中,泽泻为特征成分,当归、茯苓和芍药为主要成分,川芎

和白术在其中发挥重要作用。

6 味药对全方的贡献度结果显示当归、茯苓、芍药为本方主要组成,但这些药味也大量出现在其他的方剂配伍中,这与这些药物是中医临床常用药物的地位是相匹配的。甚至可以推论,在大多数的方剂中,诸如人参、党参、黄芪、甘草、白术、当归、茯苓、芍药等都是其主要组成,临证虽有药量加减,但药味的有无却是质的差异。因此研究丰富多彩、各具特色的方剂配伍,其药味组成的专属性或特有性应该得到更多的重视。泽泻首载于《神农本草经》,列为上品,性味甘、寒,归肾、膀胱经,功能利小便、清湿热,用于小便不利、水肿胀满、泄泻尿少、痰饮眩晕、热淋涩痛;高血脂症^[3]。《神农本草经》云“泽泻……养五脏,益气力,久服耳目聪明,不饥,延年轻身……”,但后世医家对此多有疑问,如《本草纲目》云:“神农书列泽泻于上品,复云久服轻身……愚窃疑之……。”对于泽泻,仲景认为:“水蓄渴烦,小便不利,或吐或泻,……方用泽泻,故知其用长于行水……。”统观《伤寒论》和《金匱要略》,含有泽泻的方剂共 7 首:泽泻汤、五苓散、猪苓汤、茯苓泽泻汤、牡蛎泽泻散、肾气丸和当归芍药散。前 5 方都是利水行水之名方,肾气丸去附子桂枝即后世名方六味地黄丸,唯独当归芍药散独属于《妇人病篇》下,且泽泻用量宏大。其实在《金匱》原文中也可发现,说到“妇人怀妊腹痛”在当归芍药散之前便有胶艾汤可主之,说到“妇人疝痛”后见有当归生姜羊肉汤,说到“妇人腹中诸疾痛”在当归芍药散后紧接“小建中汤也可主之”。鉴于此,我们揣测仲圣立方原旨在于,当归芍药散之证当为妇人血水之病且必见有水病,泽泻一味不可少。数据挖掘结果也初步证明了这一观点。

当然,研究结果还需要临床实践和实验研究来检验,同时进一步的数据挖掘研究还必须将药量和主治结合进来,并需要更多的样本数据,以便从更多的维度和广度来研究中药复方这一复杂体系。

参考文献:

[1] 范 明,孟小峰.数据挖掘概念与技术[M].北京:机械工业出版社,2005:3—4.

[2] 国家药典委员会.中华人民共和国药典:一部[M].北京:化学工业出版社,2005:158.

(收稿日期:2009-01-15)