

基于数据挖掘探讨慢性萎缩性胃炎的核心症状及证候要素分布规律

陈泽慧 邢恩龙 魏玥 王瑞瑞 杨晋翔

【摘要】 目的 探讨慢性萎缩性胃炎核心症状及证候要素分布规律,为阐明本病关键病机提供参考。**方法** 临床采集慢性萎缩性胃炎患者中医四诊信息、胃镜及病理检查报告构建慢性萎缩性胃炎数据库;采用隐结构模型、关联规则算法等数据挖掘方法分析慢性萎缩性胃炎核心症状及证候要素分布规律。**结果** 慢性萎缩性胃炎数据库中共包括四诊信息条目 100 个,其中出现频率为 10% 以上的条目共 59 个;通过关联规则算法得到的慢性萎缩性胃炎的核心四诊信息包括胃刺痛、口臭、纳差、咽部异物感等共 12 个;通过隐结构模型构建共挖掘慢性萎缩性胃炎主要证候要素包括气虚、湿、血瘀、热、气滞、阳虚共六种。**结论** 基于数据挖掘方法探讨慢性萎缩性胃炎核心症状及证候要素分布规律研究可进一步为本病的中医临床立法用药提供参考。

【关键词】 慢性萎缩性胃炎; 核心症状; 证候要素; 数据挖掘; 隐结构模型; 关联规则

【中图分类号】 R259 **【文献标识码】** A doi: 10.3969/j.issn.1674-1749.2022.01.006

Research on the distribution of core symptoms and syndrome elements of chronic atrophic gastritis based on data mining

CHEN Zehui, XING Enlong, WEI Yue, WANG Ruirui, YANG Jinxiang

Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China

Corresponding author: WEI Yue, E-mail: 123109668@qq.com

【Abstract】 Objective To explore the distribution of core symptoms and syndrome elements of chronic atrophic gastritis and provide references for elucidating the key pathogenesis of this disease. **Methods** The clinical data from four diagnostic methods of TCM, gastroscopy and pathological examination of patients with chronic atrophic gastritis were collected to construct a database of chronic atrophic gastritis. The methods of data mining such as hidden structure model and association rule algorithm were used to analyze the distribution of core symptoms and syndrome elements of chronic atrophic gastritis. **Results** The database of chronic atrophic gastritis includes a total of 100 items of four diagnosis information. The number of frequency, more than 10% were 59 items. The four core diagnosis information of chronic atrophic gastritis obtained by association rule algorithm including stomach-tingling, halitosis, poor appetite, and foreign body sensation in pharynx. The main syndromes of chronic atrophic gastritis were explored through the hidden structure model, including *qi*-deficiency, dampness, blood stasis, heat, *qi*-stagnation and *yang*-deficiency. **Conclusion** The study on the distribution of core symptoms and syndrome factors of chronic atrophic gas-

基金项目: 北京中医药大学基本科研业务项目(2019-JYB-XS-248)

作者单位: 100700 北京,中国中医科学院中医药信息研究所科教处(陈泽慧);首都医科大学附属北京康复医院中医康复中心(邢恩龙);北京中医药大学第三附属医院科教处(魏玥、杨晋翔);北京市丰台区中医医院 南苑医院脑病科(王瑞瑞)

作者简介: 陈泽慧(1990-),博士,助理研究员。研究方向: 中医药防治脾胃病。E-mail: chenzh@mail.citcm.ac.cn

通信作者: 魏玥(1983-),博士,副主任医师。研究方向: 中医药防治脾胃病。E-mail: 123109668@qq.com

tritis were based on data mining method, which can further provide references for the clinical legislation of TCM medication for chronic atrophic gastritis.

【Key words】 Chronic atrophic gastritis; Core symptoms; Syndrome elements; Data mining; Hidden structure model; Association rule algorithm

慢性萎缩性胃炎 (chronic atrophic gastritis, CAG) 是以固有腺体减少并常伴肠上皮化生、异型增生为病理特点的常见消化系统疾病,属于胃癌发展阶段的癌前疾病。CAG 临床症状多无特异性,亦可无明显症状,有症状者多表现为上腹部不适、饱胀、疼痛、食欲不振、反酸、嗝气等非特异性消化道症状^[1],现代医学的药物治疗多以针对性改善不同临床症状为主,对于无症状而言的患者则在治疗上存在一定局限性。通过对大量临床与基础研究的不断挖掘,越来越多的证据表明中医药在慢性萎缩性胃炎治疗方面具有良好疗效,能够减轻胃黏膜炎症反应,增强抗损伤及修复能力,甚至逆转肠化以及异型增生^[2]。辨证论治体现了中医诊疗的主要特色,通过对临床症状、舌脉等四诊信息提取后总结出的证候学类型,可反映出疾病的中医病机特点。CAG 作为脾胃系统疑难疾病之一,对其中医病机的把握有助于在无症状情况下亦可指导临床遣方用药,因此本研究基于临床采集 CAG 患者四诊信息,运用数据挖掘技术分析本病核心症状群,构建四诊信息间复杂关联网络,阐释症状—证候要素—病机之间的关键规律,从而为 CAG 临床病机理论提供依据,兹介绍如下。

1 资料与方法

1.1 研究对象

2017 年 9 月至 2019 年 10 月于北京中医药大学第三附属医院门诊、北京中医药大学东方医院消化内镜中心就诊的慢性萎缩性胃炎患者。本研究共纳入合格病例 201 例,其中男性 81 例,女性 120 例,性别比 1:1.48。年龄最小者 31 岁,年龄最大者 84 岁,平均年龄 (60.78±10.90) 岁。

1.2 诊断标准

参照 2012 年中华医学会消化病学分会于上海召开的全国慢性胃炎诊治共识会议中提出的共识意见^[3],慢性萎缩性胃炎诊断有内镜诊断和病理诊断,确诊应以病理诊断为依据。

1.3 纳入、排除与剔除标准

纳入标准: (1) 符合慢性萎缩性胃炎诊断标准,病历资料及诊断报告详实; (2) 年龄 18 岁以上患

者; (3) 患者知情同意,自愿加入本研究。

排除标准: (1) 合并消化道溃疡、胃肠道手术史、上消化道肿瘤、腹部手术等器质性病变者; (2) 合并心、脑、肝、肾等严重系统疾病; (3) 有精神系统疾病及精神病史; (4) 妊娠及哺乳期妇女。

剔除标准: (1) 不符合纳入标准而被误纳入的病例; (2) 组织病理学报告资料不全者,予以剔除; (3) 受试者未按研究要求回答或填写临床观察表,导致无效问卷者,予以剔除。

1.4 临床观察表建立

基于前期临床观察研究及相关文献,参照《中药新药临床研究指导原则 (2002 年版——试行)》^[4]、《中医临床常见症状术语规范》^[5]、《慢性萎缩性胃炎中医诊疗共识意见》^[6] 综合制定慢性萎缩性胃炎患者临床观察表,其具体内容包括: 一般资料 (姓名、年龄、性别)、病历信息 (症状、体征、病程、既往史与家族史、中医舌脉象)、辅助检查 (病理及胃镜结果)、西医诊断、中医诊断。

1.5 调查方法与质量控制

采用临床流行病学横断面研究方法,符合慢性萎缩性胃炎诊断标准的患者在取得知情同意后现场采集病史及四诊信息资料并予回收临床观察表后纳入数据统计。其中视为回收合格的临床观察表标准如下: (1) 临床观察表中所纳入符合本人情况的项目必须填写; (2) 慢性萎缩性胃炎疾病资料完备、四诊信息资料填写无缺失; (3) 所有患者的胃镜及病理检查报告应取得知情同意后拍照存档,或根据患者意愿详细记录报告信息来代替拍照存档。团队组在启动临床研究前均完成临床试验标准操作规程 (SOP) 培训,明确临床研究方案及人员分工,由至少两名以上副高级专家进行全程监督指导,严格按照纳入、排除标准对病例进行筛选,选择完全符合标准的病人入组,针对每一位患者临床调查至少由两名专业研究生同时进行信息采集,信息采集后依据剔除标准完善临床数据质量把控,以确保资料收集过程的真实可靠性,减少测量性偏倚。

1.6 统计学处理与数据分析

将所得全部临床资料录入 Microsoft Excel 2013 建立数据库,重复检查排除录入错误。应用 SPSS

23.0 软件进行频数统计,采用卡方检验分析不同证素间差异性,其中 $P<0.05$ 为存在差异, $P<0.01$ 为存在显著差异。应用 SPSS Modeler 18.0 软件进行关联规则 Apriori 算法绘制四诊信息核心组合网络图。应用 Lantern5.0 对慢性萎缩性胃炎数据库中症状数据进行隐结构分析,构建隐结构模型,并据此提取证候要素因子、归纳证候要素。

2 结果

2.1 四诊信息分布

将慢性萎缩性胃炎患者全部四诊信息进行频数频率统计,结果显示建立的慢性萎缩性胃炎患者数据库中共包括四诊信息条目 100 个,其中出现频率为 10%以上的条目共 59 个,具体条目列举见表 1。

表 1 慢性萎缩性胃炎患者四诊信息频次统计分布 (N=201)

频次 排序	四诊信息	百分比 (%)	频次 排序	四诊信息	百分比 (%)
1	苔白	70.6	31	舌淡	26.4
2	胃胀痛/胀满	66.7	32	苔黄	25.9
3	脉弦	62.2	33	脉细	24.9
4	舌暗	60.2	34	口苦	24.4
5	苔薄	57.7	35	大便干	24.4
6	头晕	51.7	36	头痛	23.4
7	泛酸	46.3	37	舌红	23.4
8	苔腻	44.8	38	脉缓	23.4
9	腰膝酸软	44.3	39	情志发作	21.9
10	舌下络瘀	44.3	40	视物不清	21.4
11	失眠	43.5	41	苔厚	21.4
12	乏力	42.3	42	大便粘滞	20.9
13	关节痛	42.3	43	心烦易怒	20.4
14	口舌干燥	41.8	44	舌边齿痕	20.4
15	脉滑	41.8	45	胃脘隐痛	18.9
16	呃逆	40.3	46	身重	18.9
17	饮食发作	38.8	47	胃灼热	18.4
18	烧心	37.8	48	口中粘腻	18.4
19	嗳气	37.3	49	气短	17.9
20	胸闷	37.3	50	腹痛/腹胀	16.9
21	情志不舒	35.3	51	四肢发凉	14.9
22	健忘	32.3	52	烦躁发热	14.9
23	胃痛有定处	31.8	53	食欲不振	13.9
24	脉沉	30.3	54	胃脘刺痛	13.4
25	嘈杂	28.9	55	咽部异物感	13.4
26	排便不爽	28.9	56	心悸	12.9
27	大便溏稀	27.9	57	渴喜热饮	12.9
28	畏寒	27.4	58	口臭	11.9
29	胁胀	26.4	59	夜间痛甚	10.9
30	多梦	26.4			

2.2 四诊信息关联规则分析

参照表 1 统计结果,应用关联规则 Apriori 算法对上述 59 个变量间关联度进行分析测量,筛选建立符合慢性萎缩性胃炎证候学研究要求的四诊信息关联规则模型。在 Apriori 算法中,支持度与置信度是判断关联规则有效性的重要指标,其中支持度表示同时包含前项 X 与后项 Y 的事务,反映了关联规则的频度;置信度是在给定前项 X 的前提下,后项 Y 的条件概率,反映了关联规则的强度。本研究在最小支持度 20%,最小置信度 80% 设置下,获得四诊信息关联规则 90 条,据此将阈值限设置在 140~200 之间绘制出四诊信息数据隐藏的关联规则网络图。结果将置信度>85%的部分四诊信息组合关联规则列出见表 2,关联规则网络图见图 1。

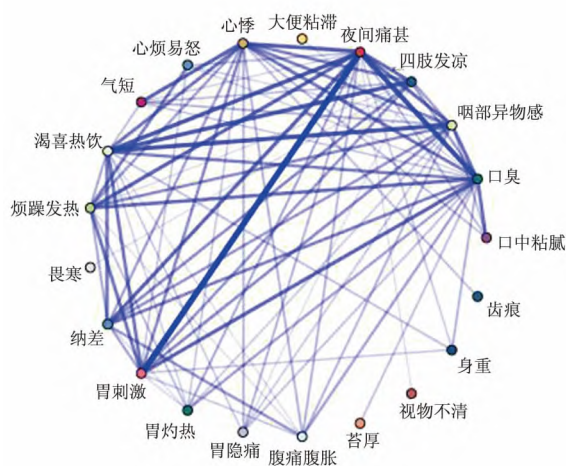


图 1 四诊信息关联规则网络图-1

为进一步明确慢性萎缩性胃炎的核心四诊信息,使在支持度 20%,置信度 80% 筛选条件下网络图更为精简,将网络图中显示连线频数值的阈值条提高,在高阈值下限时关联程度低的四诊信息不会再显示在网络图中,属性与属性之间连接线的粗细表示连接强度,线条越粗则代表两者之间关联性越强。结果显示,阈值限提高至 150~165 之间绘制出四诊信息关联规则网络图中,最终有 12 个四诊信息纳入图像;由两者之间连线粗细来判断关联强度得到胃刺痛与夜间痛甚关联最强,夜间痛甚与口臭、渴喜热饮之间关联、渴喜热饮与夜间痛甚、四肢发凉、咽部异物感之间关联较大,通过关联规则算法得到的慢性萎缩性胃炎的核心四诊信息为:胃刺痛、夜间痛甚、四肢发凉、心悸、口臭、口中粘腻、心烦易怒、纳差、烦躁发热、渴喜热饮、气短、咽部异物感。高阈值下限(150)的四诊信息关联规则网络图

见图 2。

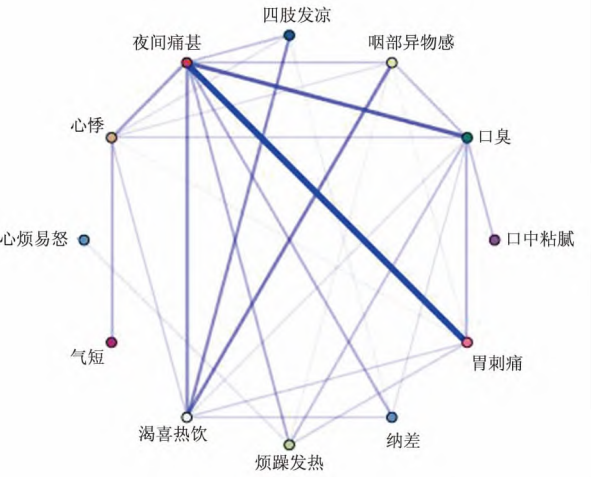


图 2 四诊信息关联规则网络图-2

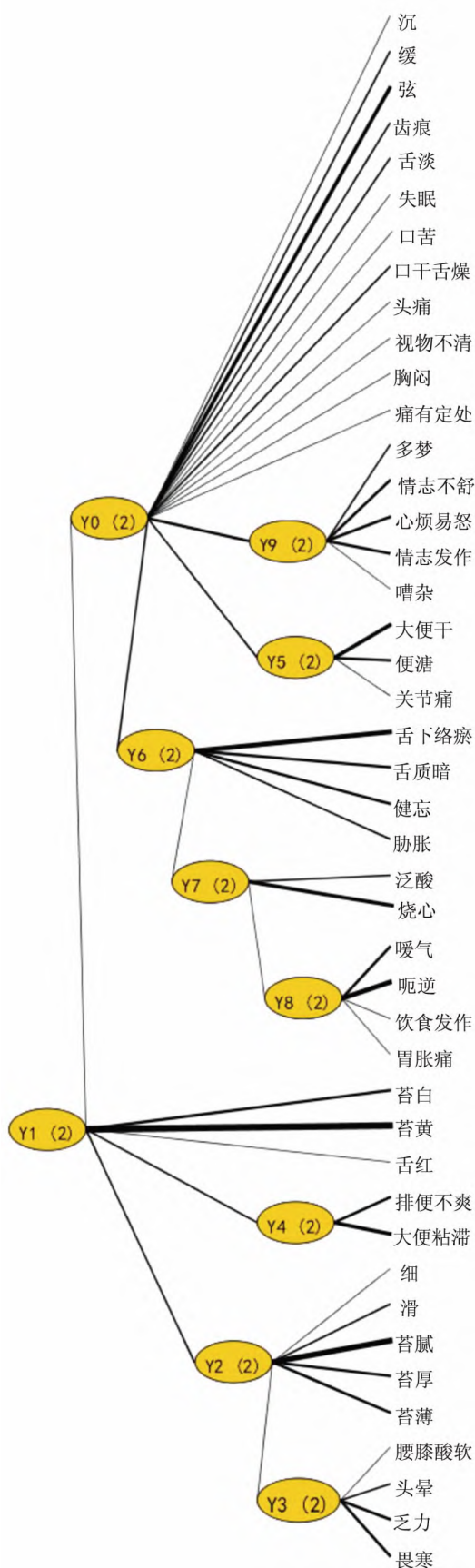
2.3 慢性萎缩性胃炎证候要素隐结构分析

参照表 1 统计结果,最终选出现频率≥20% 的 44 条四诊信息作为显变量进行爬山法 LTM-EAST 分析构建慢性萎缩性胃炎的隐结构模型,共计得到 9 个隐变量,以 Y0、Y1、Y2……Y9 表示,见图 3。其中隐变量分类个数用隐变量后括号内数字显示,2 个取值表示是/否 2 个隐类,变量间连线提示两者之间存在相关性,连线粗细代表隐变量间的互信息大小,连线越粗则两者间联系越密切。为诠释隐变量 Y 含义,Lantern 会选取若干个与隐变量 Y 相关度最高的显变量,并基于这些显变量来进行诠释。累积信息覆盖度决定了所得到的诠释结果覆

盖了多少隐变量 Y 所含的信息,累积信息覆盖度越高,选取的显变量就越多,诠释结果所覆盖的 Y 的信息也就越多。在累计信息覆盖度为 95%时,Y0 由脉弦、舌淡、脉缓、齿痕、口干舌燥、口苦、痛有定处共 7 个显变量来诠释,隐变量 Y0 把数据分为 2 个隐类,即类 Y0 = S0 和 Y0 = S1。隐类 S0 占人群总数 24%,隐类 S1 占人群总数 76%。每个显变量有两个状态,症状不出现 s0,症状出现 s1,即所在每行的 s1 列为具有该症状出现的概率。图 3 中的数据可以这样理解:隐类 S0 这类人群中具有脉弦、舌淡、脉缓、齿痕、口干舌燥、口苦、痛有定处症状的概率依次为 0、0.64、0.54、0.5、0.13、0.05、0.11;隐类 S1 这类人群中具有上述症状的概率依次为 0.82、0.14、0.14、0.11、0.51、0.31、0.39。隐类 S1 = Y0 类中,脉弦、口干舌燥、口苦、痛有定处的频率较高,反映热的证候要素,隐类 S0 = Y0 类中,舌淡、脉缓、齿痕的频率较高,反映气虚的证候要素。然而两者中 S1 = Y0 类症状(脉弦、口干舌燥、口苦、痛有定处)在隐结构模型图与 Y0 连线较粗更为密切,因此判断 Y0 的证候要素为 S1 = 热。Y0~Y9 隐变量模型诠释表见表 3,通过隐结构模型构建共挖掘慢性萎缩性胃炎主要证候要素包括气虚(Y3)、湿(Y2、Y4、Y5)、血瘀(Y6)、热(Y0、Y1)、气滞(Y8、Y9)、阳虚(Y3)共六种;由隐结构模型图观察隐变量之间连线程度得到多种证候要素之间,热(Y0)与湿(Y5)、血瘀(Y6)、气滞(Y9)较为密切。

表 2 四诊信息关联规则节选表(支持度 20%,置信度 80%)

序号	关联规则	置信度(%)	序号	关联规则	置信度(%)
1	苔腻->苔厚	97.63	17	舌暗->舌下络瘀,胃胀,苔白	88.63
2	苔白->苔薄,舌暗	94.59	18	舌暗->舌下络瘀,苔白,脉弦	88.46
3	苔白->苔薄,舌暗,脉弦	93.87	19	胃胀->呃逆,泛酸	88.37
4	苔白->苔薄,舌暗,头晕	93.33	20	苔白->苔薄,泛酸	88.23
5	苔白->苔薄,舌暗,舌下络瘀	93.18	21	舌暗->舌下络瘀,苔薄,苔白	87.23
6	舌暗->舌下络瘀,头晕,苔白	92.85	22	苔白->苔薄,头晕	86.88
7	舌暗->舌下络瘀,头晕,胃胀	92.68	23	舌暗->舌下络瘀,苔薄	86.76
8	苔白->苔薄,舌暗,胃胀	91.48	24	舌暗->舌下络瘀,脉弦	86.58
9	舌暗->健忘,苔白	91.3	25	苔白->苔薄,饮食发作	86.36
10	舌暗->舌下络瘀,关节痛	90.90	26	舌暗->舌下络瘀,健忘	85.71
11	苔白->苔薄,舌下络瘀,脉弦	90.47	27	苔白->苔薄,腰膝酸软,胃胀	85.71
12	苔白->苔薄,舌下络瘀	90.38	28	苔白->苔薄,胃胀	85.52
13	舌暗->舌下络瘀,胃胀	90.0	29	舌暗->舌下络瘀	85.39
14	舌暗->舌下络瘀,胃胀,脉弦	89.58	30	苔白->苔薄	85.34
15	舌暗->舌下络瘀,头晕	89.47	31	苔白->关节痛,舌暗	85.18
16	舌暗->舌下络瘀,头晕,脉弦	88.63	32	脉弦->头痛	85.10



3 讨论

中医临床四诊信息数据具有鲜明的多维度、非线性、多关系性、模糊性等特征,因此基于多重非线性模型的数据挖掘方法更加适用于中医证候学相关数据的解析^[7]。Apriori 算法是一种最有影响的挖掘布尔关联规则频繁项集的算法,通过 Apriori 算法可以得到高频四诊信息的组合,组合与组合间高度关联过程所产生的集合反映了慢性萎缩性胃炎的核心四诊信息群,从核心四诊信息中挖掘出的证候要素更易趋近于疾病关键病机。隐结构分析是对多种显变量特征进行汇总聚类后获得相应隐变量的过程,其特点是隐变量与显变量都取离散值,不必局限于变量间必须具有相关性,每个隐变量的诠释是根据显变量概率同现或概率互斥体现的。本研究中的显变量即四诊信息,隐变量即证候要素,证候要素(隐变量)是通过与之强关联的四诊信息(显变量)出现概率判断出来的,用数学统计描述后的证候要素具有定量性,实现了变量与变量间多重对应关系的多层次、多维度结果^[8]。

本研究在特定支持度与置信度设置下,应用关联规则 Apriori 算法进行分析测量,结果获得四诊信息关联规则 90 条,为进一步明确慢性萎缩性胃炎的核心四诊信息,将网络图阈值限提高后最终有 12 个四诊信息纳入图像,建立的四诊信息群包括:胃刺痛、夜间痛甚、四肢发凉、心悸、口臭、口中粘腻、心烦易怒、纳差、烦躁发热、渴喜热饮、气短、咽部异物感。核心四诊信息群的建立不同于简单的频数统计,而是将关联性较强的症状、体征组成多维、多层次空间网络,网络内的每条信息都与其他信息保持广泛密切的关联,因此四诊信息的核心组合并不完全等同于症状频数出现高低,本研究中可看到四诊信息出现频率在前 10 位的苍白、胃胀痛、脉弦、舌暗、苔薄等并未出现在核心四诊信息中,其原因在于频次高低不是决定信息间关联强度的必要条件,高频次四诊信息反映了普遍性存在规律,复杂关联网络中的四诊信息是反映关联性核心病机的高度概括。本研究结果中的核心四诊信息可涵盖以下证候要素:血瘀(胃刺痛、夜间痛甚)、热(口臭、烦躁发热、心烦易怒)、湿(口中粘腻、纳差)、气滞(气短、咽部异物感、心烦易怒)、阳虚(四肢发凉、渴喜热饮、气短)、气虚(纳差、气短),据此可以提出,针对慢性萎缩性胃炎基本证候要素气虚、血瘀、热、气

表 3 慢性萎缩性胃炎证候要素隐变量模型诠释表

隐变量	显变量	s0	s1	证候要素	隐变量	显变量	s0	s1	证候要素
Y0=S0 24%	弦脉	1	0	气虚	Y4=S1 31%	大便粘滞 排便不爽	0.33 0.33	0.67 0.67	湿
	舌淡	0.36	0.64		Y5=S0 57%	大便干 便溏 关节痛	1 0.51 0.71	0 0.49 0.29	
	脉缓	0.46	0.54		Y5=S1 43%	大便干 便溏 关节痛	0.43 1 0.4	0.57 0 0.6	无法判断
	齿痕	0.5	0.5		Y6=S0 48%	舌下络瘀 健忘 舌质暗	0.96 0.95 0.67	0.04 0.05 0.33	
	口干舌燥	0.87	0.13		Y6=S1 52%	舌下络瘀 健忘 舌质暗	0.18 0.42 0.14	0.82 0.58 0.86	血瘀
	口苦	0.95	0.05		Y7=S0 67%	烧心 泛酸	0.88 0.71	0.12 0.29	
	痛有定处	0.89	0.11		Y7=S1 33%	烧心 泛酸	0.11 0.19	0.89 0.81	无法判断
Y0=S1 76%	弦脉	0.18	0.82	热	Y8=S0 68%	呃逆 暖气	0.87 0.83	0.13 0.17	
	舌淡	0.86	0.14		Y8=S1 32%	呃逆 暖气	0 0.2	1 0.8	气滞
	脉缓	0.86	0.14		Y9=S0 57%	心烦易怒 情志不舒 情志发作 多梦	1 0.88 0.97 0.87	0 0.12 0.03 0.13	
	齿痕	0.89	0.11		Y9=S1 43%	心烦易怒 情志不舒 情志发作 多梦	0.52 0.34 0.52 0.55	0.48 0.66 0.48 0.45	气滞
	口干舌燥	0.49	0.51						
	口苦	0.69	0.31						
	痛有定处	0.61	0.39						
Y1=S0 74%	苔黄	1	0	热					
Y1=S1 26%	苔黄	0.01	0.99	热					
Y2=S0 61%	苔腻	0.9	0.1	无法判断					
	苔厚	0.99	0.01						
	苔薄	0.19	0.81						
Y2=S1 39%	苔腻	0	1	湿					
	苔厚	0.46	0.54						
	苔薄	0.79	0.21						
Y3=S0 81%	乏力	0.71	0.29	气虚					
	畏寒	0.85	0.14						
	头晕	0.58	0.42						
Y3=S1 19%	乏力	0	1	阳虚					
	畏寒	0.18	0.82						
	头晕	0.08	0.92						
Y4=S0 69%	大便粘滞	1	0						
	排便不爽	0.88	0.12						

滞、湿、阳虚，益气、活血、清热、理气、化湿、温阳是本病的主要治法，但临床用药必应辨证论治，若阴虚显著则以滋阴立法。

此外，基于隐结构分析提取慢性萎缩性胃炎证候要素，结果共得到 9 个隐变量，最终合并归纳为气虚、湿、血瘀、热、气滞、阳虚共 6 种。隐结构分析是在设置累计信息覆盖度为 95% 时计算机自动生成与隐变量最为相关的显变量组合，属于多维非线性客观描述。值得注意的是，隐结构分析所揭示的证候要素与核心四诊信息涵盖的证候要素组成基本吻合，并未明确显示阴虚证候要素特点，其原因可

能与本研究样本量过少、计算分析时过度筛选低频症状造成结果存在一定偏倚有关。然而有研究表明在发展为慢性萎缩性胃炎的过程中，阴虚是至为关键的病理环节之一^[9-10]。阴虚本证多因热病之后，或杂病日久伤耗阴液所致，可与气虚、血虚、阳虚、热邪等证候同时存在，或互为因果^[11]，可能基于数据挖掘在证候要素提取中并未凸显出其典型特点。不可否认的是，CAG 的发生与各种原因所致阴精耗损密切相关，经言“阳化气，阴成形”，阴精为物质形成的基础；阴精匮乏，腺体失养，体态不充，最终导致胃黏膜萎缩的形成。因此本研究认为，阴虚

也是慢性萎缩性胃炎证候因素的组成部分,扶正补虚要考虑益胃滋阴之法,特别是症见消瘦乏力、口干舌燥、舌红裂纹、脉数等类似气虚与热并见之象,实为阴虚表现时可运用如西洋参、沙参、石斛、人参叶等甘凉之品滋阴、清热、益气,实现补而不滞,泻而不虚的效果。

综上所述,本研究运用隐结构模型、关联规则算法等数据挖掘方法对 CAG 患者核心症状进行分析,得到包括气虚、湿、血瘀等在内的 CAG 证候要素六种,其所体现出的本虚标实、寒热错杂的病机特点与目前共识所述是一致的^[1]。然而某单一或较少的数据挖掘方法可能会因样本量过少、变量间缺乏相关性、过度筛选低频率症状、庞大计算量等问题造成研究结果失准^[12],这也是无监督数据挖掘所带来的局限性。在本研究中,阴虚证候要素的不凸显也可能与上述因素相关。因此,为了使研究结果尽可能地全面、客观、准确,需要以后研究考虑多数数据挖掘方法联合应用,以达多角度、多元化、择优借鉴的方式探讨 CAG 相关规律,使得数据挖掘在中医理论、实践研究中发挥更大的客观性与科学性。

参 考 文 献

- [1] 李军祥,陈諝,吕宾,等.慢性萎缩性胃炎中西医结合诊疗共识意见(2017 年) [J].中国中西医结合消化杂志,2018,26(2):121-131.
- [2] 宋青,刘震,黄达.中医药治疗慢性萎缩性胃炎的研究进展 [J].首都医科大学学报,2019,40(3):479-482.
- [3] 中华医学会消化病分会.中国慢性胃炎共识意见(2012 年,上海) [J].中华消化杂志,2013,33(1):5-12.
- [4] 中华人民共和国卫生部.中药新药临床研究指导原则(2002 年版——试行) [M].北京:中国医药科技出版社,2002:124-125.
- [5] 黎敬波,马力.中医临床常见症状术语规范 [M].北京:中国医药科技出版社,2015:1-41,75.
- [6] 唐旭东,王萍,李振华.慢性萎缩性胃炎中医诊疗共识意见 [C]//中华中医药学会脾胃病分会第二十三次全国脾胃病学术交流会论文汇编.中华中医药学会脾胃病分会,中华中医药学会,2011:7.
- [7] 张寅.慢性萎缩性胃炎证候分布特点与基于细胞动力学表型的调气活血解毒法治疗慢性萎缩性胃炎机制研究 [D].北京:北京中医药大学,2015.
- [8] 张连文,袁世宏.隐结构模型与中医辨证研究(Ⅰ)——隐结构法的基本思想及隐结构分析工具 [J].北京中医药大学学报,2006,29(6):365-369.
- [9] 张芬,王敬民,刘持年,等.应用脾阴学说治疗慢性萎缩性胃炎的浅析 [J].成都中医药大学学报,2018,41(2):91-93.
- [10] 周丽平,沈洪.从胃阴虚论治慢性萎缩性胃炎探讨 [J].山东中医杂志,2014,33(5):349-350.
- [11] 季绍良,成肇智.中医诊断学 [M].北京:人民卫生出版社,2002:105.
- [12] 李延龙,窦智丽,吴秀艳,等.数据分析方法在证候要素研究中的应用概况 [J].中华中医药杂志,2019,34(9):4175-4178.

(收稿日期:2021-04-18)

(本文编辑:蒲晓田)