

DOI:10.13288/j.11-2166/r.2025.07.010

基于潜在类别结合隐结构模型的慢性阻塞性肺疾病 肺气虚证分级量化诊断研究

李伟珂^{1,2}, 伊明洋^{1,2}, 倪园园^{1,2}, 闫丽祯^{1,2}, 关建新^{1,2}, 王世豪^{1,2}, 王慧杰¹, 王至婉^{1,2}✉1. 河南中医药大学第一附属医院, 河南省郑州市金水区人民路19号, 45000; 2. 河南中医药大学
第一临床医学院/呼吸疾病中医药防治省部共建协同创新中心/河南省中医药防治呼吸病重点实验室

[摘要] **目的** 基于潜在类别结合隐结构模型明确慢性阻塞性肺疾病(COPD)肺气虚证的分级量化诊断特征。**方法** 采集745例COPD肺气虚证患者中医四诊信息等临床资料,应用R 4.1.2软件建立潜在类别模型,通过概率参数化、参数估计与模型拟合等过程明确每位患者所属的轻、中、重度三类肺气虚证程度类别,构建轻、中、重度不同分级肺气虚证数据库。在此基础上采用Lantern 5.0软件,构建轻、中、重度不同分级肺气虚证隐结构模型,通过综合聚类并建立辨证规则。**结果** 选择频率>10%的28个症状、体征构建潜在类别模型,综合中医理论以及模型简洁程度等因素,最终判定潜在类别数为3时为最优模型,根据潜类别模型的条件概率将肺气虚证划分为轻度(298例)、中度(164例)、重度(283例)。分别构建轻、中、重度肺气虚证隐结构模型并进行综合聚类,建立辨证规则;对轻、中度肺气虚证辨证规则的共性症状比较,各症状的辨证赋值及诊断权重差异均无统计学意义($P>0.05$),因此将轻、中度肺气虚证患者作为整体进行分析,构建轻中度肺气虚证辨证规则。轻中度肺气虚证量化诊断赋值与权重:气短(赋值、诊断权重:9.3, 86.92%)、动则喘甚(8.2, 76.64%)、少气懒言(6.7, 62.62%)、纳呆(4.0, 37.38%)、便溏(4.0, 37.38%)、咳声无力(2.9, 27.10%)、喘息(2.3, 21.50%)、乏力(1.8, 16.82%)、自汗(1.7, 15.89%)、易感冒(1.6, 14.95%)、舌体胖(1.4, 13.08%)、舌边有齿痕(1.2, 11.21%)、脉沉(1.6, 14.95%),诊断阈值为10.3。重度肺气虚证量化诊断:咳声无力(15.1, 61.13%)、腰膝酸软(12.6, 51.01%)、气短(11.1, 44.94%)、少气懒言(8.3, 33.60%)、夜尿多(6.1, 24.70%)、自汗(3.7, 14.98%)、易感冒(3.5, 14.17%)、舌边有齿痕(7.8, 31.58%)、舌质淡(1.9, 7.69%)、舌苔白(5.5, 22.27%)、脉细(1.5, 6.07%),诊断阈值为23.7。**结论** 采用潜在类别结合隐结构模型明确了肺气虚证分级量化的诊断特征,可为其他基础证候量化诊断研究提供借鉴。

[关键词] 慢性阻塞性肺疾病;肺气虚证;分级量化诊断;潜在类别;隐结构模型

辨证论治是中医诊疗的核心,辨证准确与否直接影响临床疗效。传统辨证因受医生主观思维、个人经验等因素的影响,存在主观性、模糊性等弊端^[1]。证候分级量化能在一定程度上减少证候辨识的主观性,具有层次逻辑清晰、诊断条目明确等优势,有助于促进临床精准辨证^[2]。慢性阻塞性肺疾病(chronic obstructive pulmonary disease, COPD)是

一种异质性的肺部疾病,是呼吸系统常见病、多发病^[3]。肺气虚证是COPD的临床常见基础证候^[4],故以COPD肺气虚证为例开展基础证候分级量化研究更具有代表性。

中医辨证是多维度且同时聚类的过程,潜在类别模型综合了结构方程与对数线性模型的思想,其“降维简化”的方法可通过最少的潜在分类来解释显在变量间的关联程度,可无监督地刻画证候的潜在分类特征^[5]。隐结构模型是一种基于贝叶斯网络及概率图模型的无监督学习方法,通过对临床数据

基金项目:国家自然科学基金(81973791);河南省中医药学科领军人才项目(豫卫中医函[2021]8号)

✉ 通讯作者:zhiwan_w@163.com

的多维聚类划分可进行证候分级量化,在证候分级量化研究中具有重要意义^[6]。目前潜在类别结合隐结构模型在中医证候领域得到了初步应用^[7],两者结合能更准确地判断患者的证候属性及特征,尤适用于证候量化研究。因此,本研究采用潜在类别结合隐结构模型明确 COPD 肺气虚证不同分级量化诊断特征,同时为相关研究提供借鉴。本研究通过河南中医药大学第一附属医院伦理委员会审查批准(批件号 2018HL-084-01)。

1 资料与方法

1.1 病例来源

本研究采用横断面临床调查研究设计,收集 2019 年 5 月 19 日至 2022 年 12 月 31 日在全国 10 家三级甲等医院(河南中医药大学第一附属医院、河南省人民医院、河南中医药大学第二附属医院、中日友好医院、江西中医药大学附属医院、武汉市第一医院、山东中医药大学第二附属医院、内蒙古医科大学附属医院、河北省中医院、山西中西医结合医院)就诊的门诊或住院 COPD 肺气虚证患者病历资料。

1.2 西医诊断标准与中医辨证标准

西医诊断标准参照 2018 版慢性阻塞性肺疾病全球倡议(GOLD)^[8];中医辨证标准参照《慢性阻塞性肺疾病中医证候诊断标准(2011 版)》^[4]中肺气虚证辨证标准。

1.3 纳入标准

1)符合上述诊断标准与辨证标准;2)年龄 40~80 岁;3)患者签署知情同意书。

1.4 排除标准

1)神志不清、痴呆等无法正常进行沟通者;2)合并纽约心功能(NYHA)Ⅳ级^[8]、血流动力学不稳定者;3)合并活动性肺结核、气胸、胸腔积液、恶性肿瘤等影响呼吸运动功能的神经肌肉疾病;4)合并严重肝肾疾病(包括肝硬化、门脉高压及肾功能衰竭需要透析与肾移植者);5)患有其他威胁生命的严重疾病,预期生存时间<6 个月者;6)妊娠及哺乳期妇女。

1.5 调查方法

1.5.1 制定调查表 根据课题组前期研究结果^[4]制定“COPD 基础证候量化诊断临床研究调查表”。考虑到临床证候的复杂性,设立封闭式调查问卷,当出现未列出的证候类型等内容时,可由问卷研究人员或被研究者如实填写。

1.5.2 调查内容 1)人口学资料(年龄、性别等)、病程、吸烟史、合并症等;2)中医四诊信息,包括中医症状、体征、舌象、脉象等。

1.6 质量控制

制订统一的临床调查研究相关规范文件;调研前对所有研究人员进行统一培训,并进行一致性检验($Kappa \geq 0.75$,表明一致性良好);成立质量控制小组,对病例进行抽样和复核,以保证数据质量。

1.7 数据库建立及数据分析

应用 IBM SPSS Statistics 23.0 统计软件建立 COPD 肺气虚证临床调查数据库。

1.7.1 一般资料分析 采用 IBM SPSS Statistics 23.0 统计软件,定量资料符合正态分布者采用均数 $\pm$ 标准差($\bar{x} \pm s$)进行描述;不符合正态分布者采用中位数(四分位数)[$M(P_{25}, P_{75})$]进行描述。定性资料采用频次、频率进行描述。

1.7.2 潜在类别模型 使用 R 4.1.2 软件 poLCA 软件包进行潜在类别建立模型。具体过程:1)模型参数化:主要包括潜在类别概率和条件概率。2)参数估计、模型拟合与评价:参数估计一般选用最大似然法,其常用的迭代算法为期望最大化(EM)算法;模型拟合检验方法为似然比卡方(G^2)、Pearson χ^2 ;模型评价:一般认为赤池信息准则(AIC)、贝叶斯信息准则(BIC)、似然比卡方统计量(ΔG^2)、对数似然数(log-likelihood)各数值均达到最小,模型适配较好,其潜在类别模型最优。3)潜在分类:通过贝叶斯公式计算得到最优模型个体的后验类别属性。4)结果解释:对潜在类别下条目的条件概率特点进行潜类别解释。通过上述步骤,明确肺气虚证轻、中、重不同严重程度类别;并分别构建数据库。

1.7.3 隐结构模型 采用 Lantern 5.0 软件(香港科技大学研制),分别建立轻、中、重度肺气虚证隐结构模型。具体过程包括:采用双步隐树分析(LTM-EAST)算法分析不同分级肺气虚证症状,构建隐结构模型,通过结构图、互信息曲线图、隐类概率分布表并结合中医专业理论知识对隐变量及隐类进行诠释并综合聚类,采用 BIC 评分评价模型,由 Lantern 5.0 软件导出不同分级肺气虚证的辨证规则并对症状进行赋值、建立阈值。

2 结果

2.1 一般资料

共采集 750 例 COPD 肺气虚证患者临床资料,

将基线资料重点信息缺失的病例视为无效病例，共筛选出 745 例有效病例纳入结果分析。其中男性 582 例（78.12%）、女性 163 例（21.88%）；平均年龄为（67.13±9.09）岁；前一年急性加重次数 1（1，2）次；前一年感冒次数 2（0，3）次；病程 5（2，10）年。

2.2 患者症状分布情况

对 745 例患者症状进行统计分析，其中咳嗽出现频率最高（61.89%），其次为气短（61.32%）、舌质淡（61.12%）、咳痰（52.41%）等，出现频率>20%的症状详见表 1。

表 1 745 例慢性阻塞性肺疾病肺气虚证患者出现频率>20%的症状分布情况

Table 1 Symptom distribution with a frequency larger than 20% in 745 patients with chronic obstructive pulmonary disease and lung qi deficiency syndrome					
症状	频次	频率/%	症状	频次	频率/%
咳嗽	461	61.89	咳声无力	290	38.99
气短	457	61.32	舌苔薄	280	37.60
舌质淡	455	61.12	少气懒言	266	35.71
咳痰	390	52.41	余沥不尽	255	34.17
喘息	373	50.05	肢体困倦	237	31.76
乏力	369	49.59	腰膝酸软	213	28.64
舌苔白	368	49.39	脉沉	212	28.48
神疲	362	48.57	脉弦	210	28.13
胸闷	337	45.24	纳呆	192	25.82
动则喘甚	337	45.24	耳鸣	187	25.05
易感冒	333	44.67	舌体胖	186	24.95
自汗	329	44.21	腹胀	184	24.64
恶风	324	43.29	夜尿多	163	21.88
脉细	300	40.22	大便溏	160	21.41

2.3 潜在类别分析结果

2.3.1 模型拟合与选择 为选择最优的肺气虚证

表 3 慢性阻塞性肺疾病肺气虚证不同潜在类别概率及条件概率（条件概率>0.35）

Table 3 Probabilities and conditional probabilities (greater than 0.35) of different latent classes in lung qi deficiency syndrome of chronic obstructive pulmonary disease		
潜在类别数	潜在类别概率	中医症状（条件概率）
类别 1	0.3822	气短(0.9957)、动则喘甚(0.9654)、咳嗽(0.6575)、喘息(0.6374)、神疲(0.6261)、咳声无力(0.5985)、乏力(0.5867)、咳痰(0.5533)、少气懒言(0.5440)、舌质淡(0.5156)、胸闷(0.5153)、易感冒(0.4919)、余沥不尽(0.4767)、腰膝酸软(0.4591)、自汗(0.4560)、恶风(0.4528)、舌苔白(0.4034)、脉细(0.3815)
类别 2	0.3955	舌质淡(0.5923)、咳嗽(0.5852)、舌苔白(0.5587)、咳痰(0.4954)、舌苔薄(0.4565)、脉细(0.4482)、乏力(0.4143)、恶风(0.4140)、喘息(0.4116)、易感冒(0.3988)、胸闷(0.3947)、自汗(0.3927)、气短(0.3649)、神疲(0.3600)
类别 3	0.2223	舌质淡(0.8058)、纳呆(0.7219)、便溏(0.7120)、腹胀(0.6149)、咳嗽(0.6098)、气短(0.5511)、舌苔白(0.5382)、咳痰(0.5228)、舌体胖(0.5200)、自汗(0.5021)、乏力(0.4793)、神疲(0.4590)、舌边有齿痕(0.4579)、易感冒(0.4508)、恶风(0.4400)、喘息(0.4179)、胸闷(0.4426)、脉细(0.3594)

潜在类别模型，本研究依次构建类别数为 1-5 的潜在类别模型并进行比较。表 2 示，AIC、BIC、 ΔG^2 、对数似然数从类别 1 开始逐渐降低，至第 3 个潜在类别时拟合优度各指标均为最小，从第 4 个潜在类别开始升高，第 5 个潜在类别又开始下降。因此，当潜在类别个数为 3 时，模型拟合优度各项指标均为最小，综合中医理论以及模型简洁程度等因素，最终判定潜在类别数为 3 时为最优模型。

表 2 慢性阻塞性肺疾病肺气虚证潜在类别模型的拟合优度
Table 2 Goodness-of-fit of the latent class model for lung qi deficiency syndrome in chronic obstructive pulmonary disease

潜在类别数	对数似然数	BIC	AIC	ΔG^2
类别 1	-34721.86	69655.86	69499.71	39957.17
类别 2	-33709.05	67849.96	67532.09	37931.55
类别 3	-33103.44	66858.48	66378.89	36720.34
类别 4	-34721.19	70974.30	69688.38	39939.20
类别 5	-34415.84	69998.48	69139.68	39328.51

注：BIC，贝叶斯信息准则；AIC，赤池信息准则； ΔG^2 ，似然比卡方统计量。

2.3.2 模型参数估计 运用 EM 算法对参数进行估计。表 3 示，轻、中、重度肺气虚证人群的潜在类别概率分别为 0.3822、0.3955、0.2223，提示从 745 例患者中随机抽取一个病例属于类别 1 的概率为 0.3822，属于类别 2 的概率为 0.3955，属于类别 3 的概率为 0.2223。条件概率为不同潜类别下各症状发生的概率，如在类别 1 肺气虚证患者中气短条件概率最高为 0.9957，其次是动则喘甚(0.9654)、咳嗽(0.6575)等；3 个潜在类别的条件概率分布图见附图 1（扫描本文二维码获取）。表 3 显示类别 1 中各症状条件概率最大，为重度肺气虚证人群；其次为类别 3 为中度肺气虚证人群；类别 2 最小，为轻度肺气虚证人群。

2.3.3 后验类别概率 后验概率分析可将每个患者分类, 根据 Bayes 公式可算出个体取值的后验概率, 概率最大者则属于该类别。按照 745 例患者入组先后列举前 4 位及后 4 位患者的各类别的后验概率 (见表 4), 如 001 号患者, 其类别 3 的后验概率最高为 0.723218, 则该患者归属于类别 3, 即中度肺气虚人群。最终 745 例 COPD 患者分类为轻度肺气虚证 298 例, 中度肺气虚证 164 例, 重度肺气虚证 283 例。

表 4 745 例慢性阻塞性肺疾病肺气虚证患者不同严重程度类别潜在分类示例

Table 4 Examples of latent classification for different severity categories in 745 patients with chronic obstructive pulmonary disease and lung qi deficiency syndrome

患者编号	后验概率			类别
	类别 1	类别 2	类别 3	
001	0.267215	0.009567	0.723218	3
002	0.140860	0.136494	0.722646	3
003	0.950754	0.047758	0.001488	1
004	0.837513	0.073702	0.088785	1
.....				
742	0.841378	0.047724	0.110898	1
743	0.808018	0.155967	0.036015	1
744	0.046433	0.080216	0.873351	3
745	0.099279	0.899612	0.001109	2

2.4 隐结构分析结果

2.4.1 不同分级肺气虚证隐结构模型及综合聚类 构建轻、中、重度不同分级肺气虚证隐结构模型并进行综合聚类。以轻度肺气虚证为例, 将所有隐变量中提取到与轻度肺气虚证密切相关的隐变量进行综合聚类, 综合分析隐变量的结构图、互信息曲线图、隐类概率分布表及显变量的模型参数, 并导出辨证规则。

2.4.2 模型诠释 以 Z1 (轻度肺气虚证) 为例, 各症状对应的互信息、信息覆盖度、条件概率等见

表 5 与附图 2 (扫描本文二维码获取)。综合聚类把轻度肺气虚证分为两类, 即每个显变量有两种状态, 症状不出现 S0, 症状出现为 S1, Z=S1 为有轻度肺气虚证的患者, 此类患者占轻度肺气虚证患者的 37%。Z1=S1 这类患者症状出现概率较高, 可认为有轻度肺气虚证。条件概率则可以通过该组症状所表现出的变化来定量刻画所代表的证候特征。少气懒言的条件概率最高为 0.81, 易感冒为 0.37、神疲为 0.55。肺气虚证各综合聚类模型及信息曲线见附图 2-附图 4 (扫描本文二维码获取)。

2.4.3 不同分级肺气虚证辨证规则的建立 综合分析结构图、互信息曲线图、隐类概率分布及显变量的模型参数, Lantern5.0 软件导出不同分级肺气虚证的辨证规则, 对不同分级肺气虚证的主要症状进行赋值、建立阈值, 并计算其诊断权重, 详见表 6-表 8。

表 6 慢性阻塞性肺疾病轻度肺气虚证辨证规则的诊断权重

Table 6 Diagnostic weights of syndrome differentiation rules for mild lung qi deficiency syndrome in chronic obstructive pulmonary disease

症状	赋值	诊断权重 (%)
气短	5.5	76.39
动则喘甚	3.8	52.78
少气懒言	3.6	50.00
神疲	2.9	40.28
易感冒	2.6	36.11
自汗	2.0	27.78
乏力	1.2	16.67
脉细	0.4	6.94
脉沉	0.4	5.56
恶风	0.4	5.56
阈值	7.2	-

2.4.4 肺气虚证辨识规范的建立 研究发现轻、中度肺气虚证辨证中共性症状较多, 且其赋值及诊断权重相近, 为进一步评价轻、中度肺气虚证辨识

表 5 慢性阻塞性肺疾病肺气虚证综合聚类模型及症状条件概率分布

Table 5 Conditional probability distribution of symptoms in the comprehensive clustering model for lung qi deficiency syndrome in chronic obstructive pulmonary disease

综合聚类模型 (Z=S1)	BIC 评分	显变量 (症状的条件概率) =S1
轻度肺气虚证 (0.37)	-2916.84	少气懒言 (0.81)、易感冒 (0.37)、神疲 (0.55)、自汗 (0.59)、恶风 (0.59)、喘息 (0.68)、舌质淡 (0.64)
中度肺气虚证 (0.63)	-2015.06	舌苔白 (0.72)、舌质淡 (0.93)、神疲 (0.52)、动则喘甚 (0.52)、自汗 (0.63)、易感冒 (0.58)、脉细 (0.88)
重度肺气虚证 (0.46)	-1097.46	舌边有齿痕 (0.66)、咳声无力 (0.78)、舌质淡 (0.51)、少气懒言 (0.48)、夜尿多 (0.56)、舌苔白 (0.73)、腰膝酸软 (0.61)

注: BIC, 贝叶斯信息准则。

表 7 慢性阻塞性肺疾病中度肺气虚证辨证规则的诊断权重

Table 7 Diagnostic weights of syndrome differentiation rules for moderate lung *qi* deficiency syndrome in chronic obstructive pulmonary disease

症状	赋值	诊断权重 (%)
气短	8.4	67.74
动则喘甚	5.8	46.77
自汗	4.5	36.29
神疲	4.1	33.06
舌苔白	3.8	30.65
舌质淡	3.8	30.65
易感冒	3.5	28.23
脉沉	2.5	20.16
乏力	2.0	16.13
咳声无力	1.2	9.68
喘息	1.0	8.06
恶风	0.8	6.45
脉细	0.5	4.03
阈值	12.4	—

表 8 慢性阻塞性肺疾病重度肺气虚证辨证规则的诊断权重

Table 8 Diagnostic weights of syndrome differentiation rules for severe lung *qi* deficiency syndrome in chronic obstructive pulmonary disease

症状	赋值	诊断权重 (%)
咳声无力	15.1	61.13
腰膝酸软	12.6	51.01
气短	11.1	44.94
少气懒言	8.3	33.60
夜尿多	6.1	24.70
自汗	3.7	14.98
易感冒	3.5	14.17
舌边有齿痕	7.8	31.58
舌质淡	1.9	7.69
舌苔白	5.5	22.27
脉细	1.5	6.07
阈值	24.7	—

模型效果,本研究通过对共性症状诊断权重值比较,发现气短、动则喘甚、神疲、易感冒、自汗、咳嗽、乏力、咳痰各症状在轻、中度肺气虚证辨证中差异均无统计学意义($P>0.05$),详见表9,提示两者的辨识区分程度一般。因此,结合临床实际,将轻、中度肺气虚证患者作为整体进行分析,进一步优化不同分级证候量化辨识模型。构建轻中度肺气虚证隐结构模型,共得到14个隐变量,通过综合隐变量的结构图、互信息曲线图、隐类概率分布表及显变量的模型参数,由Lantern5.0软件导出轻中度肺气虚证的辨证规则,对诠释轻中度肺气

虚证的主要症状进行赋值、建立阈值,并计算其诊断权重,详见表10与附图5(扫描本文二维码获取)。

表 9 慢性阻塞性肺疾病轻、中度肺气虚证共性症状诊断权重比较

Table 9 Comparison of diagnostic weights for common symptoms in mild and moderate lung *qi* deficiency syndrome in chronic obstructive pulmonary disease (%)

证候	气短	动则喘甚	神疲	易感冒	自汗	咳嗽	乏力	咳痰
轻度	76.39	52.78	40.28	36.11	27.78	20.83	16.67	13.89
中度	67.94	46.77	33.06	28.23	36.29	28.23	16.13	24.19
<i>P</i> 值	0.071	0.396	0.304	0.225	0.225	0.250	0.849	0.225
χ^2 值	3.249	0.720	1.057	1.471	1.471	1.325	0.036	1.471

表 10 慢性阻塞性肺疾病轻中度肺气虚证辨证规则

Table 10 Syndrome differentiation rules for mild and moderate lung *qi* deficiency syndrome in chronic obstructive pulmonary disease

症状	赋值	诊断权重 (%)
气短	9.3	86.29
动则喘甚	8.2	76.64
少气懒言	6.7	62.62
便溏	4.0	37.38
纳呆	4.0	37.78
咳声无力	2.9	27.10
喘息	2.3	21.50
乏力	1.8	16.82
自汗	1.7	15.89
易感冒	1.6	14.95
舌体胖	1.4	13.08
舌边有齿痕	1.2	11.21
脉沉	1.6	14.95
阈值	10.3	—

3 讨论

早在《黄帝内经》就提出了量度的概念^[9],体现了中医学定量辨证的内核^[10]。中医证候量化由来已久,基于传统的模糊定量辨证思想,目前广泛开展了证候量化研究,使原有的定性诊断变成定性与定量相结合的计量诊断,对证候量化诊断研究具有更深层次的意义,但仍存在以下问题:1)证候量化主要围绕症状等的量化,较少涉及证候或基础证候的量化分级;2)证候量化研究中多为症状“轻、中、重”等级的划分,未区分症状对证候贡献度的差异;3)研究方法具有局限性,传统统计学方法难以有效反映证候量化中各关键因素间的复杂、非线性、模糊关系。因此,深入研究基础证候的分级量化,对把握辨证的准确性和客观性、加强辨证的规范化、提高临床诊疗水平具有重要的现实意义。

无监督算法不依赖训练标签,仅根据数据本身

便可提取数据集中的隐藏信息,可精准捕捉数据的潜在规律和特征,适宜于探讨证候内部不同分级的本质特性^[10]。潜在类别与隐结构模型均为无监督算法,两者结合在证候领域得到了一定的应用,证实了两种方法的有效性及其合理性,但将其应用于证候分级量化尚属空白。潜在类别模型“降维简化”的思想^[11],可有效处理分类资料,无监督地判定不同分级证候的人群潜在类别,达到证候分级,但无法有效识别证候内部各症状特征的重要程度。隐结构模型可客观、定量地揭示证候与症状的关系,但对证候严重程度的识别精度欠佳。因此,本研究将两者联合应用于证候分级量化研究中,以弥补单一方法的不足,为其他相关研究提供方法学示范。

本研究选取了频率大于 20% 的肺气虚证 28 个症状构建潜在类别模型,基于模型的个体观测反应模式,识别了不同分级的肺气虚证人群的症状特征。与其他类别相比,类别 1 中各症状的条件概率最高,如气短在类别 1 中条件概率为 0.9957,类别 2 中为 0.3649,类别 3 中为 0.5511,由此我们发现同一症状在不同人群中严重程度不同。因此,潜在类别模型可用来探索证候内部不能直接观察到的异质性,以刻画证候的多维性,实现了分级诊断。最终 745 例 COPD 患者中轻度肺气虚证 298 例,中度肺气虚证 164 例,重度肺气虚 283 例。

本研究在潜在类别模型的基础上,通过隐结构模型揭示了症状特征与证候不同分级间的复杂关系,且放宽了局部独立假设,可更全面地反映症状间的关联。研究建立了轻、中、重度不同分级患者的隐结构模型,并明确了不同分级肺气虚证的辨证规则;其中轻、中度肺气虚证共性症状较多,且证候贡献程度均无统计学差异,提示两者的辨识区分度一般。考虑为研究中仅通过各四诊信息条目的“有”或“无”来判别,导致模型难以仅通过症状有无将两者有效区分。结合临床实际,将轻、中度肺气虚证患者作为整体进行分析,建立了轻中度肺气虚证量化辨识模型。研究进一步证实了隐结构模型在一定程度上对证候的轻重程度有明确的量化认识。

本研究建立的肺气虚证量化辨识规范中不同症状有不同的赋值及诊断权重,不同分级的肺气虚证有不同的诊断阈值。研究中发现,证候表现越严重、其诊断阈值越高,如轻中度肺气虚证诊断阈值为 10.3,重度肺气虚证诊断阈值为 23.7;同一症状在不同分级证候中的赋值不同,对辨识该证候的贡献度也不同,如气短在轻中度肺气虚证中赋值为

9.3,诊断权重为 86.92%,而在重度肺气虚证中赋值为 11.1,诊断权重为 44.94%。此外,轻中度肺气虚证辨识规范中存在纳呆、便溏、舌边有齿痕等一般认为是脾气虚的表现;考虑为肺气虚损,子病及母,致脾气受损,进而肺脾俱虚,因此临床中 COPD 稳定期补虚药中以补气药居多,其功效兼补肺脾^[12],这与《慢性阻塞性肺疾病中医证候诊断标准(2011 版)》^[4]中 COPD 气虚证涉及脾脏,表现为肺脾气虚证一致。重度肺气虚证中存在腰膝酸软、夜尿多,此类症状一般认为是肾气虚证的典型表现,考虑为肺气虚日久,肺气虚程度愈深,病情反复发作,迁延不愈,肺为气之主,肾为气之根,由肺及肾,肾气必虚,肾失摄纳,而致肺肾气虚,结果符合临床实际。也进一步证实了潜在类别和隐结构模型两种无监督算法在基础证候分级量化辨识规范建立研究中的适宜性。

综上所述,本研究采用潜在类别结合隐结构模型建立了轻中度及重度肺气虚证量化诊断辨识规范。在疾病框架下,基础证候有不同分级,不同分级下诊断指标各有权重;量化分级明确,可重复性强,具有一定创新性。但尚存在不足之处:一是四诊信息的采集均为二分类,未进行分级量化,忽略了其严重程度对证候分级量化的贡献差异,致使模型的预测精准度有待提高;二是在对数据的处理方面,本研究基于无监督方法虽在一定程度上减少了主观因素的影响,但本质还是属于症状的聚类,缺乏评价的客观标准,后续将针对此开展进一步的研究。

参考文献

- [1] 杨涛,朱学芳. 中医辨证智能化研究现状及发展趋势[J]. 南京中医药大学学报, 2021, 37(4): 597-601.
- [2] 李蒙,张培彤,郭秀伟. 中医证候分级量化诊断标准的研究进展[J]. 中华中医药杂志, 2017, 32(8): 3641-3643.
- [3] Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease 2025 report [EB/OL]. (2024-11-11) [2025-01-10]. <https://gold-copd.org/2025-gold-report/>. 3.
- [4] 中华中医药学会内科分会肺系病专业委员会. 慢性阻塞性肺疾病中医证候诊断标准(2011 版)[J]. 中医杂志, 2012, 53(1): 80-84.
- [5] 王琪,胡良平,高颖,等. 潜在类别分析在缺血性中风病患者证候数据中的应用[J]. 中国卫生统计, 2013, 30(5): 638-640.
- [6] 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.
- [7] 李敏, 何庆勇, 刘旭东, 等. 基于潜在类别结合隐结构分析的血脂异常患者中医常见证候特征研究[J]. 世界科学技术-中医药现代化, 2022, 24(5): 2102-2113.
- [8] Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease 2018 report [EB/OL]. (2017-11-15) [2024-07-24]. <https://gold-copd.org/2018-gold-report/>.
- [9] 刘玉良. 《黄帝内经》量值理论和定量辨证思想初探[J]. 中医杂志, 2012, 53(22): 1895-1898.
- [10] 方金海. 基于无监督学习的糖尿病数据异常检测方法研究[D]. 南京: 南京邮电大学, 2022.
- [11] 张恒艳, 叶建州, 周也, 等. 潜类别结构方程在中医证候演变分析中的运用概述[J]. 时珍国医国药, 2018, 29(10): 2476-2477.
- [12] 杨勤军, 何振豪, 史友, 等. 基于症状-证候-药物的新安医家治疗肺胀医案数据挖掘研究[J]. 北京中医药大学学报, 2022, 45(6): 594-602.

Study on Graded Quantitative Diagnosis of Lung *Qi* Deficiency Syndrome in Chronic Obstructive Pulmonary Disease Based on Latent Class Analysis Combined with Hidden Structure Model

LI Weike^{1,2}, YI Mingyang^{1,2}, NI Yuanyuan^{1,2}, YAN Lizhen^{1,2}, GUAN Jianxin^{1,2}, WANG Shihao^{1,2}, WANG Huijie¹, WANG Zhiwan^{1,2}

1. The First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou, 45000; 2. The First Clinical Medical College of Henan University of Chinese Medicine/Collaborative Innovation Center for Chinese Medicine and Respiratory Diseases co-constructed by Henan province & Education Ministry of P. R. China/Henan Key Laboratory of Chinese Medicine for Respiratory Disease

ABSTRACT Objective To clarify the graded quantitative diagnostic characteristics of lung *qi* deficiency syndrome in chronic obstructive pulmonary disease (COPD) based on latent class analysis combined with a hidden structure model. **Methods** Clinical data, including the four diagnostic methods of traditional Chinese medicine (TCM), were collected from 745 COPD patients with lung *qi* deficiency syndrome. Latent class modeling was performed using R 4.1.2 software, and each patient was classified into one of three severity categories (mild, moderate, or severe) based on probabilistic parameterization, parameter estimation, and model fitting. A database was established for different severity levels of lung *qi* deficiency syndrome. Based on this, Lantern 5.0 software was used to construct hidden structure models for mild, moderate, and severe lung *qi* deficiency syndrome, and syndrome differentiation rules were developed through comprehensive clustering. **Results** A latent class model was constructed using 28 symptoms and signs with a frequency greater than 10%. Considering TCM theory and model simplicity, the optimal model was determined when the number of latent classes was three, categorizing lung *qi* deficiency syndrome into mild (298 cases), moderate (164 cases), and severe (283 cases). Hidden structure models were separately developed for each severity level, and syndrome differentiation rules were established. A comparison of common symptoms in the syndrome differentiation rules for mild and moderate lung *qi* deficiency syndrome showed no statistically significant differences in diagnostic values and weights ($P > 0.05$), leading to their combined analysis and the development of a unified syndrome differentiation rule. Value and weight of quantitative diagnosis of mild-to-moderate lung *qi* deficiency syndrome were as followed: shortness of breath (diagnostic value 9.3, diagnostic weight 86.92%), dyspnea on exertion (8.2, 76.64%), low voice and reluctance to speak (6.7, 62.62%), poor appetite (4.0, 37.38%), loose stools (4.0, 37.38%), weak cough sound (2.9, 27.10%), wheezing (2.3, 21.50%), fatigue (1.8, 16.82%), spontaneous sweating (1.7, 15.89%), susceptibility to colds (1.6, 14.95%), swollen tongue (1.4, 13.08%), teeth marks on the tongue edge (1.2, 11.21%), deep pulse (1.6, 14.95%), with a diagnostic threshold of 10.3. Value and weight of quantitative diagnosis of severe lung *qi* deficiency syndrome were as followed: weak cough sound (15.1, 61.13%), soreness and weakness of the waist and knees (12.6, 51.01%), shortness of breath (11.1, 44.94%), low voice and reluctance to speak (8.3, 33.60%), frequent nocturia (6.1, 24.70%), spontaneous sweating (3.7, 14.98%), susceptibility to colds (3.5, 14.17%), teeth marks on the tongue edge (7.8, 31.58%), pale tongue body (1.9, 7.69%), white tongue coating (5.5, 22.27%), thin pulse (1.5, 6.07%), with a diagnostic threshold of 23.7. **Conclusion** The combination of latent class analysis and a hidden structure model effectively clarified the graded quantitative diagnostic characteristics of lung *qi* deficiency syndrome, providing a reference for the quantitative diagnosis of other fundamental syndromes in TCM.

Keywords chronic obstructive pulmonary disease; lung *qi* deficiency syndrome; quantitative grading diagnosis; latent class analysis; hidden structure model

(收稿日期: 2024-08-09; 修回日期: 2025-01-15)

[编辑: 柴倩云]