

· 数据挖掘 ·

基于隐结构结合 Logistic 回归分析探讨 9 323 例古籍咳嗽医案证候分布

陈丽平^{1,2}, 李建生^{2*}, 杨淑慧¹, 卞华¹, 庞立业¹

(1. 南阳理工学院 河南省张仲景方药与免疫调节重点实验室, 河南 南阳 473004;

2. 河南中医药大学 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, 郑州 450046)

[摘要] 目的:探索中医古籍咳嗽医案证候分布规律。方法:提取古籍中医肺病数据库内咳嗽病案 9 323 例,用 Lantern 5.0 (孔明灯)软件构建前 50 位症状和中药共 100 个显变量隐结构模型,对不同隐节点进行诠释,根据隐结构综合聚类权重量化揭示症状和证候、中药(方剂)和证候间的辨证规则,通过二分类 Logistic 回归分析拟合不同病性和证候间的相关关系。结果:9 323 例咳嗽医案涉及证候 204 种,频次>100 的证候有 18 种。将前 50 位症状和 50 味中药建模得到 35 个隐变量,98 个隐类,10 个综合聚类模型,其中 Z5 的阈值 6.7 最高,Z6 咯痰信息覆盖度 52% 最高,Z7 的分值 19 最高。二分类 Logistic 回归模型拟合出不同病性和 5 类证型的相关关系,其中外风和风热犯肺证的优势比最高达 88.919,痰热蕴肺证与病性热、痰的优势比为 51.594 和 15.861,痰湿阻肺证与湿、痰、饮的优势比分别为 31.415,34.370 和 4.936。从证候的频数分布、以症推证、以药(方)测证得出咳嗽的常见证候有 14 种,分别为外感咳嗽有风寒袭肺、风热犯肺、风燥犯肺,内伤咳嗽有痰湿阻肺、痰热壅肺、肝火犯肺、肺阴亏虚、肺气阴两虚、脾肺气虚、肺肾阴虚、肺热阴虚、外寒内饮、脾胃虚弱、痰瘀阻肺。结论:新发现的咳嗽证候有肺热阴虚、外寒内饮、脾胃虚弱、痰瘀阻肺证,临证多以复合证候为主,如肺气阴两虚、肺肾阴虚。中医证候的辨证带有一定的主观性,权值的高低提示不同症状对证候的贡献度有差异,对证候推断有一定指导意义,隐结构模型结合 Logistic 回归分析在一定程度上能解决中医辨证的定量问题,可用于疾病证候分布的挖掘。

[关键词] 隐结构模型; Logistic 回归分析; 咳嗽; 证候

[中图分类号] R249;R289;R22;R2-031;R33 **[文献标识码]** A **[文章编号]** 1005-9903(2021)14-0175-08

[doi] 10.13422/j.cnki.syfjx.20210815

[网络出版地址] <https://kns.cnki.net/kcms/detail/11.3495.R.20210421.1551.004.html>

[网络出版日期] 2021-04-21 17:33

Syndrome Distribution of 9 323 Cough Cases in Ancient Chinese Medical Books Based on Latent Structure Model and Logistic Regression Analysis

CHEN Li-ping^{1,2}, LI Jian-sheng^{2*}, YANG Shu-hui¹, BIAN Hua¹, PANG Li-ye¹

(1. Henan Key Laboratory of Zhang Zhongjing Formulae and Herbs for Immune Regulation, Nanyang Institute of Technology, Nanyang 473004, China;

2. Respiratory Disease Prevention and Control of Provincial Department of Traditional Chinese Medicine to Build Collaborative Innovation Center, Henan Key Laboratory of Chinese Medicine for Respiratory Disease, Henan University of Chinese Medicine, Zhengzhou 450046, China)

[Abstract] **Objective:** To explore the syndrome distribution of cough cases in ancient Chinese medical books. **Method:** A total of 9 323 cough cases in the database of lung diseases in ancient Chinese medical books

[收稿日期] 20200820(017)

[基金项目] 国家自然科学基金项目(81704200);河南省中医药科学研究专项(2019ZY3033)

[第一作者] 陈丽平,博士,副教授,从事中医文献整理与挖掘利用研究,Tel:0377-83995705,E-mail:942202160@qq.com

[通信作者] *李建生,博士,教授,从事呼吸疾病中医药防治的临床与基础研究,Tel:0377-65676568,E-mail:li_js8@163.com

were extracted. Lantern 5.0 was used to construct the latent structure model for the 100 manifest variables based on the first 50 symptoms and 50 Chinese herbal medicines, and different latent nodes were interpreted. The syndrome differentiation patterns of syndromes with symptoms and Chinese herbal medicine (formula) were quantitatively revealed by the comprehensive clustering weights of latent structure. The correlation of diseases with syndromes was fitted through the binary Logistic regression analysis. **Result:** There were 204 syndromes involved in 9 323 cough cases with 18 syndromes showing a frequency higher than 100. As demonstrated by the model established on the first 50 symptoms and 50 Chinese herbal medicines, 35 latent variables, 98 latent classes, and 10 comprehensive clustering models were obtained, where Z5 was the highest in the threshold value (6.7), Z6 in the information coverage of productive cough (52%), and Z7 in the score (19). The binary Logistic regression model fitted the correlation between different disease types and five syndromes, where the dominance ratio of external wind to the syndrome of wind-heat invading lung reaching up to 88.919, those of syndrome of phlegm-heat accumulating in lung to diseased heat and sputum 51.594 and 15.861, and those of the syndrome of phlegm-dampness obstructing lung to dampness, phlegm, and fluid retention 31.415, 34.370, and 4.936, respectively. **Conclusion:** The newly discovered cough syndromes included lung heat and yin deficiency, external cold and internal fluid retention, weakness of spleen and stomach, and phlegm and blood stasis in lung. In most cases, multiple syndromes were observed clinically, such as syndrome of deficiency of both Qi and Yin in lung combined with yin deficiency in lung and kidney. Since differentiation of traditional Chinese medicine (TCM) syndrome is subjective, the weight can indicate the difference in the contributions of different symptoms to the syndrome, which is of guiding significance for syndrome inference. The latent structure model combined with Logistic regression analysis can solve the problem of quantification in TCM syndrome differentiation and can be used to explore the syndrome distribution of diseases.

[Keywords] latent structure model; Logistic regression analysis; cough; syndrome

咳嗽是呼吸系统常见症状或疾病之一,其常见的证候有外感风寒、外感风热、燥邪犯肺、痰湿犯肺、痰热郁肺、肝火犯肺、肺阴虚证^[1]。证候是中医学辨证论治的依据与精髓,其综合概括了机体在内外因素作用下的整体性反应状态^[2],临床上表现为症状、体征、舌象、脉象等四诊信息的综合,通过不同的组合和关联构成不同的证和病^[3]。中医证候研究是中医药现代化的关键环节,在中医理论体系中占据核心地位,是连接基础理论和临床辨治的纽带。文献研究是证候研究的基础,通过文献研究系统挖掘证候类别及其与证素间的关系,探索证候分布规律与筛选证候诊断指标,为临床研究提供依据^[4]。

中医古籍医案中对咳嗽证候记载较多,与现代认识有不同之处。以往针对证候的研究主要采用因子分析、聚类分析^[5]找出症状的集合,或者根据证素(如病性和病位)间组合频数挖掘出证候分布,具有一定的意义,但对证候的推断不够全面。笔者依据隐结构的辨证规则以方测证得出9 323例咳嗽医案的常见证候,包括凉燥袭肺、肾阴虚、肺肾阴虚、外寒内饮、风热犯肺、肝火犯肺、肺热炽盛、痰湿阻肺、痰瘀阻肺9种^[6]。本文采用隐结构模型的辨证

规则揭示证候分布,用 Logistic 回归分析明确不同病性证素对证候的影响度,拟通过综合运用证候频数分布、以药(方)测证、以症测证等多种方法挖掘出咳嗽的证候分布规律,使各种统计结果相互印证、补充,为完善咳嗽证候分类及辨证标准的制定提供依据。

1 资料与方法

1.1 数据来源与规范 提取古籍中医肺病数据库内咳嗽病案9 323例^[7],包括汉代至1911年间历代中医先贤编纂撰写的中医著作825种,如《外台秘要方》《济阳纲目》《赤水玄珠》《证治准绳》《叶天士医案》《临证指南医案》《王仲奇医案》《邵氏方案》《医方配本》《张聿青医案》《丹溪摘玄》《邵兰荪医案》《南雅堂医案》《清代名医医案精华》《沈菊人医案》等。根据《中医临床诊疗术语》^[8]对咳嗽的证候名称进行判定,据《证素辨证学》^[9]中有关证候要素组成原则对相关数据进行拆分,从而规范证候和证候要素的名称。据《中医药学名词》^[10]和《常用中药名与别名手册》^[11]对中药名称进行规范。将多词一义、一证(症)多名或一药多名,用一词表达;拆分组合概念,包括证候、药组等。

1.2 数据处理与分析 运用隐结构模型可以使辨证定量化,利用权重大小分析隐变量与显变量(症状或中药)的关系,进而体现症状或中药对相应证候的贡献度,为建立证候或证候要素的诊断标准提供依据^[12]。Logistic回归分析是一种非线性回归模型,通过研究自变量(常为疾病的危险因素)与因变量(疾病发生)之间的定量关系,对因变量进行分类和预测。Logistic回归公式 $Y=1/(1+e^{-w'x})$ 。其中, Y 为Logistic函数, e 为自然对数的底数, x 为独立变量, w' 为第 t 个Logistic函数相关的系数的向量。Logistic回归模型里包含下列主要统计量, B 为偏回归系数。 SE 为偏回归系数的标准误。 $Wald$ 统计量用于检验总体偏回归系数与0有无显著差异,其等于偏回归系数与标准误之商的平方。 Sig 为显著性,其 <0.05 有统计意义。 $Exp(B)$ 为优势比(比数比),即偏回归系数的反自然对数,其值越大表示因变量与自变量的关系越密切。

表1 9 323例咳嗽医案频数>100次的证候分布

Table 1 9 323 cases of cough medical case frequency>100 syndrome distribution

证候	频数/次	占比/%	证候	频数/次	占比/%	证候	频数/次	占比/%
痰热壅肺	1 450	15.55	肺热阴虚	564	6.45	脾肾阳虚	142	1.52
痰湿阻肺	1 087	11.66	风热犯肺	465	4.99	阴虚火旺	142	1.52
肺阴亏虚	1 204	12.91	肺气阴两虚	294	3.15	寒痰阻肺	118	1.27
风寒袭肺	798	8.56	肺肾阴虚	250	2.68	燥邪犯肺	113	1.21
寒饮停肺	617	6.62	肺气虚	232	2.49	风邪犯肺	111	1.19
脾肺气虚	587	6.30	外寒内饮	162	1.74	肝火犯肺	107	1.15

2.1 构建症状与中药的隐结构 针对9 323例咳嗽医案排名前50位的症状和中药,共100个显变量构建隐结构模型,见图1,用爬山法(LTM-EAST)进行分析,累积信息覆盖率为95%,共得到35个隐变量,用 $Y_0, Y_1 \dots Y_{34}$ 表示,每个隐变量下涉及2个到多个隐类,计98个隐类,其中涉及2个隐类的有21个隐变量,评价标准为贝叶斯信息测度(BIC)评分,模型分数-173297.02。图1以 Y_0, Y_5, Y_{26}, Y_{29} 为中心, Y_{26} 涉及6个隐类, $Y_0, Y_3, Y_5, Y_{27}, Y_{29}$ 涉及5个隐类, $Y_2, Y_4, Y_{28}, Y_{31}, Y_{33}, Y_{34}, Y_8$ 涉及3个隐类,综合聚类分为10组, Y_0, Y_1, Y_2, Y_5 为Z1; Y_{10}, Y_3, Y_4 为Z2; Y_{27}, Y_{28} 为Z3; $Y_{26}, Y_{29} \sim Y_{31}$ 为Z4; $Y_{29}, Y_{32} \sim Y_{34}$ 为Z5; Y_6, Y_{25}, Y_7 为Z6; $Y_{24}, Y_8 \sim Y_{10}$ 为Z7; $Y_{11}, Y_{12}, Y_{22}, Y_{23}, Y_{13}$ 为Z8; $Y_{14}, Y_{15}, Y_{20}, Y_{21}$ 为Z9; $Y_{15} \sim Y_{19}$ 为Z10,前5组以中药聚类为主,能以方测证,后5组主要为症状聚类,能以症推证。

根据综合聚类权重对证候进行做最终判定,与

本研究先用Lantern 5.0隐结构分析软件(香港科技大学研制)构建咳嗽医案症状和中药的隐结构模型,展现显变量和隐变量的关系,线条越粗显示贡献度越大^[13]。通过对不同隐节点的诠释,如显变量与综合聚类间的互信息、信息覆盖度、累积互信息、累积信息覆盖度、隐类概率等,提取证候分布规律,根据综合聚类权重量化揭示症状和证候、中药(方剂)和证候间的辨证规则,分别用阈值和分值表示。辨证规则是用来量化症状或中药在证候判别时提供的支持证据,主要指标为对数优势比(lgOR)^[6]。用SPSS 20.0软件通过二分类Logistic回归分析拟合不同病性和证候间的相关关系。

2 结果

9 323例咳嗽医案涉及证候204种,频数>100次的证候有18种,内伤咳嗽以痰热壅肺、痰湿阻肺、肺阴亏虚为、寒饮停肺证为主,外感咳嗽常见风寒袭肺、风热犯肺,见表1。

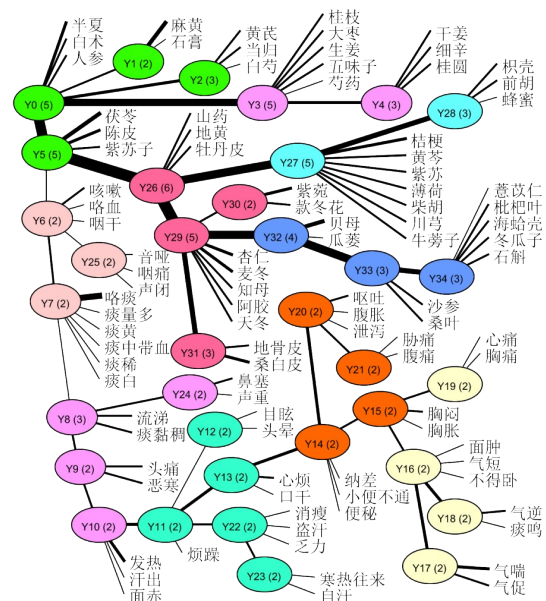


图1 9 323例咳嗽医案排名前50位中药与症状隐结构模型

Fig. 1 Top 50 Chinese herbs and symptom cryptic structure model in 9 323 cough medical cases

Z1关系密切的是半夏和茯苓,其与Z1的互信息是0.22和0.12,信息覆盖度是40%和22%,Z1聚类的患者中应用该药组的概率是62%,Z1总信息量0.55,证型分值4.2,见表2。

表2 9 323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z1权重

Table 2 Cluster Z1 weight of top 50 symptoms in 9 323 cough medical cases and hidden structure model of TCM

Z1	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	P=0.62(s0,s1)	P=0.38(s0,s1)	分值
半夏	0.22	40	0.22	40	1.00,0.00	0.43,0.57	-1
茯苓	0.12	22	0.33	60	0.92,0.08	0.70,0.30	2
白术	0.10	19	0.40	73	1.00,0.00	0.94,0.06	7
人参	0.10	18	0.46	84	0.93,0.07	0.89,0.11	2
陈皮	0.09	16	0.50	91	0.91,0.09	0.53,0.47	1
麻黄	0.05	9	0.52	96	0.97,0.03	0.74,0.26	-1

与Z2关系密切的是人参和五味子,其与Z2的互信息是0.11和0.1,信息覆盖度是29%和28%,Z2聚类的患者中应用该药组的概率是71%,Z2总信息量0.36,证型分值3.3,见表3。

表3 9 323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z2权重

Table 3 Cluster Z2 weight of top 50 symptoms in 9 323 cough medical cases and hidden structure model of TCM

Z2	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	P=0.71(s0,s1)	P=0.29(s0,s1)	分值
人参	0.11	29	0.11	29	0.95,0.05	0.56,0.44	2
五味子	0.10	28	0.19	52	0.97,0.03	0.68,0.32	0.3
半夏	0.08	22	0.24	67	0.92,0.08	0.60,0.40	-1
干姜	0.06	18	0.28	77	0.99,0.01	0.91,0.09	3.5
白术	0.06	16	0.31	85	0.99,0.01	0.78,0.22	7
细辛	0.05	14	0.32	89	0.99,0.01	0.90,0.10	-3.6
桂枝	0.03	9	0.33	91	0.98,0.02	0.94,0.06	16.6
生姜	0.03	8	0.34	95	0.98,0.02	0.85,0.15	2.4

与Z3关系密切的是桔梗,其与Z3的互信息是0.13,信息覆盖度是45%,Z3聚类患者中应用该药组概率是67%,Z3总信息量0.29,证型分值3.4,见表4。

表4 9 323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z3权重

Table 4 Cluster Z3 weights of top 50 symptoms in 9 323 cough medical cases and hidden structure model of TCM

Z3	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	P=0.67(s0,s1)	P=0.33(s0,s1)	分值
桔梗	0.13	45	0.13	45	0.94,0.06	0.42,0.58	2.7
枳壳	0.08	28	0.18	64	0.99,0.01	0.71,0.29	4.5
前胡	0.08	26	0.21	75	0.98,0.02	0.67,0.33	3
紫苏叶	0.05	18	0.23	81	0.98,0.02	0.77,0.23	2.4
薄荷	0.04	15	0.25	88	0.98,0.02	0.78,0.22	4
牛蒡子	0.03	9	0.27	94	0.99,0.01	0.86,0.14	-4.3
黄芩	0.02	7	0.28	97	0.95,0.05	0.80,0.20	4.6

与Z4关系密切的是麦冬,其与Z4的互信息是0.1,信息覆盖度是22%,Z4聚类患者中应用该药组概率是65%,Z4总信息量0.43,证型分值5,见表5。

与Z5关系密切的是浙贝母和苦杏仁,其与Z5

的互信息是0.23和0.13,信息覆盖度是43%和24%,Z5聚类的患者中应用该药组的概率是68%,Z5总信息量0.55,证型分值6.7,见表6。

与Z6关系密切的是咯痰和咳嗽,其与Z6的互信息是0.11和0.1,信息覆盖度是52%和48%,Z6聚

表5 9323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z4权重

Table 5 Cluster Z4 weight of first 50 symptoms in 9323 cough medical cases and hidden structure model of TCM

Z4	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	$P=0.65(s_0, s_1)$	$P=0.35(s_0, s_1)$	分值
麦冬	0.10	22	0.10	22	0.97, 0.03	0.88, 0.12	1
苦杏仁	0.09	21	0.18	43	0.83, 0.17	0.35, 0.65	4
桑白皮	0.09	21	0.25	58	0.97, 0.03	0.61, 0.39	5
地黄	0.07	15	0.30	70	0.98, 0.02	0.97, 0.03	-4
紫菀	0.05	11	0.34	78	0.97, 0.03	0.75, 0.25	6
天冬	0.05	11	0.37	85	1.00, 0.00	0.91, 0.09	3
阿胶	0.04	10	0.39	90	0.99, 0.01	0.89, 0.11	7
知母	0.04	9	0.42	95	0.98, 0.02	0.84, 0.16	2

表6 9323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z5权重

Table 6 Cluster Z5 weight of first 50 symptoms in 9323 cough medical cases and hidden structure model of TCM

Z5	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	$P=0.68(s_0, s_1)$	$P=0.32(s_0, s_1)$	分值
浙贝母	0.23	43	0.23	43	0.98, 0.02	0.45, 0.55	5.8
苦杏仁	0.13	24	0.32	59	0.83, 0.17	0.76, 0.24	4
桑叶	0.09	16	0.36	67	1.00, 0.00	0.93, 0.07	8.1
南沙参	0.08	15	0.42	76	1.00, 0.00	0.74, 0.26	5.8
枇杷叶	0.07	12	0.44	80	0.99, 0.01	0.94, 0.06	4.3
麦冬	0.06	12	0.48	88	0.94, 0.06	0.56, 0.44	1
冬瓜子	0.05	9	0.49	91	1.00, 0.00	0.98, 0.02	15.4
天冬	0.03	6	0.51	94	0.98, 0.02	0.80, 0.20	3
知母	0.03	6	0.53	96	0.97, 0.03	0.76, 0.24	2

类的患者中有该症状组的概率是49%, Z6总信息量 0.22, 证型分值4, 见表7。

表7 9323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z6权重

Table 7 Cluster Z6 weight of first 50 symptoms in 9323 cough medical cases and hidden structure model of TCM

Z6	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	$P=0.49(s_0, s_1)$	$P=0.51(s_0, s_1)$	分值
咯痰	0.11	52	0.11	52	0.94, 0.06	0.54, 0.46	4
咳嗽	0.10	48	0.20	92	0.40, 0.60	0.04, 0.96	4
咽干	0.01	5	0.21	96	1.00, 0.00	0.96, 0.04	6

与Z7关系密切的是鼻塞, 其与Z7的互信息是 0.1, 信息覆盖度是47%, Z7聚类的患者中有该症状组的概率是92%, Z7总信息量0.22, 证型分值6, 见表8。

表8 9323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z7权重

Table 8 Cluster Z7 weight of first 50 symptoms in 9323 cough medical cases and hidden structure model of TCM

Z7	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	$P=0.92(s_0, s_1)$	$P=0.08(s_0, s_1)$	分值
鼻塞	0.10	47	0.10	47	1.00, 0.00	0.44, 0.56	19
发热	0.07	30	0.17	77	0.93, 0.07	1.00, 0.00	-9
流涕	0.05	22	0.19	86	1.00, 0.00	0.68, 0.32	8
头痛	0.04	19	0.21	95	0.99, 0.01	0.90, 0.10	4
恶寒	0.04	17	0.22	100	1.00, 0.00	0.92, 0.08	5

与Z8关系密切的是心烦和头晕, 其与Z8的互信息均是0.02, 信息覆盖度是30%和70%, Z8聚类的患者中有该症状组的概率是97%, Z8总信息量0.07, 证型分值6, 见表9。

与Z9关系密切的是呕吐和纳差, 其与Z9的互信息是0.05和0.03, 信息覆盖度是34%和23%, Z9聚类的患者中有该症状组的概率是91%, Z9总信息量0.14, 证型分值5.7, 见表10。

表9 9 323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z8权重

Table 9 Cluster Z8 weight of first 50 symptoms in 9 323 cough medical cases and hidden structure model of TCM

Z8	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	$P=0.97(s_0, s_1)$	$P=0.03(s_0, s_1)$	分值
心烦	0.02	30	0.02	30	0.97, 0.03	0.62, 0.38	5
头晕	0.02	22	0.05	70	1.00, 0.00	0.80, 0.20	5
目眩	0.01	19	0.05	71	1.00, 0.00	0.83, 0.17	5
盗汗	0.01	15	0.06	80	0.99, 0.01	0.84, 0.16	5
消瘦	0.01	12	0.06	86	0.99, 0.01	0.85, 0.15	4
自汗	0.01	11	0.07	95	1.00, 0.00	0.88, 0.12	5

表10 9 323例咳嗽医案前50位症状与中药隐结构模型综合聚类Z9权重

Table 10 Cluster Z9 weight of first 50 symptoms in 9 323 cough medical cases and hidden structure model of TCM

Z9	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	$P=0.91(s_0, s_1)$	$P=0.09(s_0, s_1)$	分值
呕吐	0.05	34	0.05	34	0.99, 0.01	0.67, 0.33	5.1
纳差	0.03	23	0.1	73	0.99, 0.01	0.78, 0.22	5
腹胀	0.02	17	0.12	84	1.00, 0.00	0.86, 0.14	6.1
腹痛	0.02	15	0.13	91	1.00, 0.00	0.89, 0.11	7.1
泄泻	0.02	12	0.14	96	1.00, 0.00	0.89, 0.11	5.3

与Z10关系密切的是胸闷和气喘,其与Z10的互信息是0.07和0.06,信息覆盖度是31%和27%,

Z10聚类的患者中有该症状组的概率是74%,Z10总信息量0.23,证型分值3.9,见表11。

表11 9 323例咳嗽医案排名前50位症状与中药隐结构模型综合聚类Z10权重

Table 11 Cluster Z10 weight of first 50 symptoms in 9323 cough medical cases and hidden structure model of TCM

Z10	互信息	信息覆盖度/%	累积互信息	累积信息覆盖度/%	$P=0.74(s_0, s_1)$	$P=0.26(s_0, s_1)$	分值
胸闷	0.07	31	0.07	31	0.98, 0.02	0.93, 0.07	3.9
气喘	0.06	27	0.13	56	0.75, 0.25	0.37, 0.63	2.4
气逆	0.05	23	0.17	73	0.95, 0.05	0.66, 0.34	3.1
心痛	0.03	13	0.19	81	1.00, 0.00	1.00, 0.00	1.5
气短	0.02	8	0.2	87	0.99, 0.01	0.92, 0.08	3.6
痰鸣	0.02	7	0.21	91	1.00, 0.00	0.94, 0.06	5
胸痛	0.01	6	0.22	94	0.99, 0.01	0.98, 0.02	0.3
气促	0.01	6	0.23	97	0.97, 0.03	0.88, 0.12	2.1

2.2 基于 Logistic 回归分析病性与证候的相关关系 9 323 例咳嗽医案的证候经规范后有 125 种,排名前 11 位(频数>200 次,频率>2%)分别是痰热壅肺(15.55%),痰湿(浊)阻肺(11.66%),肺阴亏虚(12.91%),风寒袭肺(8.56%),寒饮停肺(6.62%),脾肺气虚(6.30%),肺热阴虚(6.05%),风热犯肺(4.99%),肺气阴两虚(3.15%),肺肾阴虚(2.68%)和肺气虚(2.49%)。将所有证型、病性赋值(1=有,0=无),以排名前 11 位证型为因变量,相关病性为自变量,采用二分类 Logistic 回归模型,以 $P \leq 0.10$ 为纳入模型标准,以 $P > 0.15$ 为剔除标准,其中的 5 类证型的拟合结果为痰热蕴肺证与病性热、痰的优势比为

51.594 和 15.861,脾肺气虚证与气虚的优势比为 11.377,肺气虚证与气虚、血虚的优势比为 66.710 和 26.306,风寒袭肺证与寒、外风的优势比为 10.749 和 10.674,风热犯肺证与外风、热的优势比为 88.919 和 5.537,痰湿阻肺证与湿、痰、饮的优势比分别为 31.415, 34.370 和 4.936,其回归方程为①痰热壅肺证 $= -2.728 + 3.943 \text{热} - 4.996 \text{燥} + 2.764 \text{痰} - 4.351 \text{湿} - 3.081 \text{饮}$;②脾肺气虚证 $= -2.573 + 2.432 \text{气虚} - 2.621 \text{寒} - 2.609 \text{外风}$;③肺气虚证 $= -3.682 + 4.2 \text{气虚} - 2.694 \text{阴虚} + 3.27 \text{血虚}$;④风寒袭肺证 $= -3.269 + 2.375 \text{寒} + 2.368 \text{外风} - 4.343 \text{痰}$;⑤风热犯肺证 $= -4.67 + 4.488 \text{外风} + 1.711 \text{热} - 6.189 \text{寒} - 3.283 \text{温} - 1.575 \text{湿}$;⑥痰湿阻

肺证 = -1.942 + 3.447 湿 - 4.344 外风 + 3.537 痰 + 1.597 饮。

3 讨论

咳嗽的证候分布在临床上主要以小样本及单一挖掘为主,孙欣强^[14]将符合纳入标准的86例感染后咳嗽辨证分为5种类型,其中以风寒束肺证及风热犯肺证为主。余银^[15]对120例小儿上气道咳嗽综合征中医证候进行挖掘得到风热犯肺、肺脾气虚、阴虚肺热、风痰犯肺4个证候。何德平等^[16]对102例慢性咳嗽医案聚类分析显示证候分布以风寒犯肺、肺气亏虚、阴虚肺燥、痰湿阻肺、肝火犯肺为主。这些结果一定程度上揭示了咳嗽的证候分布规律,但不够系统和完善。中医证候的量化是信息的表达,因权值高低的差异,带有一定主观性,量化时主要取决于不同症状对证候的贡献度,需要采用多种挖掘方法实现优势互补。证候具有多维性和复杂性的特点,部分古籍咳嗽医案明确记载其证候类型,经规范后可直接统计证候的频数。还有部分医案证候字段是缺如的,需根据中医理论从其他方面进行推测,如以药(方)测证、以症测证等。不同病性对同一证候的相关性也有差异,Logistic回归分析能解决不同病性和证候间的相关关系大小。下面从证候频数、以药(方)测证、以症测证分析咳嗽证候分布特点,基于Logistic回归分析结果探讨病性和证候的关系。

3.1 证候频数 表1资料显示古籍医案中咳嗽证候出现的频率依次是痰热壅肺(1450次,15.55%),痰湿阻肺(1087次,11.66%),肺阴亏虚(1204次,12.91%),风寒袭肺(798次,8.56%),寒饮停肺(617次,6.62%)。其中有3类证候均和痰饮水湿有关,痰和热是居首位的内因和病性,二者相合干肺。痰(湿)为阴邪,其性重浊黏滞,痰饮水湿等日久易郁而化热;热是阳邪,其性炎上急迫,痰(湿)与热相合,导致痰热壅肺证出现频率最高。痰、热、风是咳嗽的高频病因,痰热壅肺和痰湿阻肺是咳嗽最常见的证候,与有关医家认识一致^[17-18]。

3.2 以药(方)测证 图1中综合聚类Z1~Z5的流线较粗,均为中药的综合聚类,提示隐结构揭示的证候分型较好,通过中药组方的疗效进行以理明证和以法使方,将证的理论通过所用方药进行反向推测来反证辨证的合理性,以药(方)测证是探索中医证候诊断的有效方法^[19]。Z1~Z5经隐类权重计算后表2~6提示Z1含二陈汤,Z2含小青龙汤,Z3含银翘散,Z4清燥救肺汤,Z5含南沙参麦冬汤的主要药组,以

药(方)测证得出Z1~Z5对应的证型有痰湿阻肺、外寒内饮、风热犯肺、肺热阴虚、肺阴虚。根据隐结构的辨证规则,以Z5为例说明基于中药隐结构综合聚类后进行以药(方)测证的相对合理性,其他证型亦可采用此方法进行判定。通过对数优势比公式可以得到Z5各味中药的分值浙贝母(5.8),苦杏仁(4),桑叶(8.1),南沙参(5.8)是正面辨证依据,肺阴虚证的阈值为6.7。将辨证规则应用于Z5,若 $5.8+4+8.1+5.8+4.3+1+15.4+3+2>6.7$,则判定为肺阴虚证。

3.3 以症推证 图1中综合聚类Z6~Z10的流线较细,均为症状的综合聚类,提示隐结构揭示的证候分型相对较差,需要结合专家辨证对证候推断的结果进行遴选,不能单纯依靠数据统计的直观结果。症状是疾病反映于外的现象,它是判断病种、辨别证型的主要依据,综合聚类能将反映同一证候不同侧面的隐节点综合,其权重大小为以症测证提供辨证依据。以图1中综合聚类Z6和Z7为例,Z6由Y6咳嗽、咯血、咽干,Y7咯痰、痰量多、痰黄、痰中带血,Y25咽痛、音哑组成,多因痰热之邪随肺气上逆而致咳嗽、咯大量黄稠痰,热盛伤津动血则咯血、咽干、痰中带血。表7中咯痰对痰热壅肺证的信息覆盖度 $52%>50%$,信息覆盖度主要用来量化一组症状对证的信息表达程度,考察与某证相关的症状中有几个症状可完全概括此证的基本特征,即完成“辨证”的过程^[20]。累积信息覆盖度解决纳入多少显变量能描述某一证型的分布特征,咯痰、咳嗽、咽干3个症状的累积信息覆盖度达 $96%>95%$,以症测证得出Z6为痰热壅肺证。Z7由Y24鼻塞、声重,Y8~Y10流涕、头痛、恶寒、发热等组成,《诸病源候论》曰“咳嗽者,肺感于寒,微则成咳嗽也,气虚为微寒客皮毛,入伤于肺则不足,成咳嗽。此由邪热客于肺也。上焦有热,其人必饮水,水停心下,则肺为之浮,肺主于咳,水气乘之,故咳嗽。”^[21]表8中的症状当纳入到恶寒时,其累积信息覆盖度达 $100%>95%$,即可剔除其他冗余数据。将辨证规则应用于Z7,若 $19-9+8+4+5>6$,则判定为风寒袭肺证。Z8的总信息量0.02最低,提示这组症状对证候判定时拟合较差,结合表1证候频数分布中肺气阴两虚294例,予以保留。Z9为脾胃虚弱证,与Z9关系密切的是呕吐和纳差,体现在临证中慢性咳嗽培土生金之法。Z10气郁伤肺为咳嗽并发症常见的证候,经临床专家审核和临床不太符合,Z10证候的辨证规则虽提示正值,但是与临床不符,予以剔除。

3.4 基于Logistic回归模型判定病性与证候的相关性 Logistic回归分析中与痰热壅肺证关系最密切的病性是病性热优势比(51.594)>痰(15.861),提示热的显著性和优势性明显高于痰。与肺气虚关系最密切的病性是气虚优势比(66.710),其次是血虚(26.306),因气为血之帅,血为气之母,二者互生互化、互根互用。病性风优势比(10.674)≈寒(10.749)和风寒袭肺证的相关性相当,病性寒略高于外风,和肺的特性喜温恶寒有关。与风热犯肺证关系最密切的病性是外风优势比(88.919),其次是热(5.537),提示外风的显著性和优势性明显高于热,因风邪为六淫之首,其他外邪多随风邪侵袭人体,所以外感咳嗽常以风为先导,或夹寒,或夹热,或夹燥,表现为风寒、风热、风燥等。病性痰优势比(31.415)和湿(34.370)与痰湿阻肺证的相关性相当,另外病性饮(4.936)与此证型的相关性也较大,痰饮水湿均为津液代谢异常的产物,因存在部位不同而有差异,但均贯穿于咳嗽发生发展的始终。

总之,咳嗽的常见证型有外感咳嗽,包括风寒袭肺、风热犯肺、风燥犯肺,内伤咳嗽包括痰湿阻肺、痰热壅肺、肝火犯肺、肺阴亏虚、肺气阴两虚、脾肺气虚、肺肾阴虚、肺热阴虚、外寒内饮、脾胃虚弱、痰瘀阻肺。从证候的频数分布、以药(方)测证、以症测证得出咳嗽的常见证候和第二版《实用中医内科学》^[22]不同的有肺热阴虚、外寒内饮、脾胃虚弱,临证多以复合证候为主,如肺气阴两虚、肺肾阴虚。同时结合前期的数据挖掘结果,9 323例咳嗽医案中新增证候除和本文一致的外寒内饮证,还有痰瘀阻肺证^[6],数据挖掘结果可为咳嗽证候分布的探讨提供多种方法,不能单纯根据结果下定论,还需要和临床实际相结合,才能做出有益的探索。

[利益冲突] 本文不存在任何利益冲突。

[参考文献]

[1] 郭德海,郑光,张洁,等. 基于文本挖掘的慢性咳嗽中医诊治规律研究[J]. 中国中医药信息杂志,2019,26(10):101-104.
[2] 庄丽,李运伦,朱羽硕,等. 基于量表学与组学建立中医证候宏观辨证体系的可行性探讨[J]. 中国实验方剂学杂志,2019,25(20):166-172.
[3] 李建生,余学庆,王至婉. 病证结合诊疗模式下证候诊断标准建立的关键环节[J]. 中医杂志,2013,54(15):1261-1264.
[4] 春柳,谢洋,赵虎雷,等. 基于文献的中医证候诊断规

范研制现状分析[J]. 中医杂志,2021,62(2):169-172.
[5] 杨勤军,韩明向,李泽庚,等. 基于聚类分析和因子分析的慢性咳嗽用药规律探索[J]. 中国实验方剂学杂志,2019,25(19):155-160.
[6] 陈丽平,李建生,邢静宇,等. 基于隐结构辨证规则的9 323例中医古籍咳嗽病案的方-证分析[J]. 中国实验方剂学杂志,2020,26(18):179-187.
[7] 陈丽平,李建生,蔡永敏,等. 中医肺病数据库构建及挖掘的思路与方法[J]. 中华中医药杂志,2017,32(12):5530-5534.
[8] 国家技术监督局. 中医临床诊疗术语:GB/T 167512-1997[S]. 北京:中国标准出版社,1997.
[9] 朱文锋. 证素辨证学[M]. 北京:人民卫生出版社,2008.
[10] 中医药学名词审定委员会. 中医药学名词[M]. 北京:科学出版社,2005.
[11] 谢宗万. 常用中药名与别名手册[M]. 北京:人民卫生出版社,2009.
[12] 王天芳,张连文,赵燕,等. 隐结构模型及其在中医证候研究中的应用[J]. 北京中医药大学学报,2009,32(8):519-526.
[13] 张连文,袁世宏. 隐结构模型与中医辨证研究(1)——隐结构法的基本思想以及隐结构分析工具[J]. 北京中医药大学学报,2006,29(6):365-369.
[14] 孙欣强. 感染后咳嗽中医证候分布规律的初步研究[D]. 广州:广州中医药大学,2010.
[15] 余银. 小儿上气道咳嗽综合征中医证候分布规律研究[D]. 长沙:湖南中医药大学,2015.
[16] 何德平,林琳,吴蕾,等. 慢性咳嗽中医证候分布规律探讨[J]. 广州中医药大学学报,2008,25(6):560-562.
[17] 乔羽,姜德友. 基于古今医案数据分析咳嗽证候规律[J]. 河南中医,2017,37(7):1240-1243.
[18] 刘三峰,卢峰,侯玉泉,等. 基于近10年文献分析的咳嗽病证型及证素研究[J]. 亚太传统医药,2018,14(4):101-103.
[19] 赵立凝. 慢性浅表性胃炎脾虚证诊断标准确立及以药测证研究[D]. 广州:广州中医药大学,2005.
[20] 陈丽平,蔡永敏,李建生. 基于隐结构模型对名老中医慢性支气管炎医案病机的探讨[J]. 中华中医药杂志,2017,32(2):801-806.
[21] 巢元方. 诸病源候论[M]. 北京:华夏出版社,1997:74.
[22] 王永炎,严世芸. 实用中医内科学[M]. 2版. 上海:上海科学技术出版社,2009:172-177.

[责任编辑 顾雪竹]