

基于现代名老中医临床著作的肺病 数据库建立及数据挖掘

蔡永敏 陈丽平 李建生 李庆磊 孙曙明 李成文

摘要 目的 建立现代名老中医临床著作的肺病数据库,探索现代名老中医治疗肺系疾病的证治规律。
方法 以现代临床著作中名老中医治疗肺病的临证经验和医案文献为数据来源,通过字段、程序设计,构建结构型数据库,并在此基础上采用数据挖掘的方法:频数分析、聚类分析、因子分析及关联规则对肺系疾病的证治规律进行研究。
结果 建立了具备文献检索、信息统计、数据挖掘等功能的现代名老中医临床著作肺病数据库,共录入 34 414 条数据,其中医案医话 28 045 条(81.49%),临床经验 6 369 条(18.51%);医案医话中男性 14 048 条(50.09%),女性 9 466 条(33.75%),男女之比为 1.48:1,未注名性别者(均为医话文献)4 531 条(16.16%)。可实现多字段的相互关联分析、聚类分析、因子分析、关联规则等数据挖掘。
结论 本文采集到的数据以医案医话为主,肺系疾病的患病率男性明显高于女性;本数据库可查寻各种肺系疾病相关字段的集中表现趋势;针对不同字段进行各种适应性的数据挖掘,能发现各种肺系疾病的诊治规律;可实现大样本临床文献的症、证、方、药等多变量多角度的数据挖掘。

关键词 肺系疾病;名老中医;数据挖掘;临床著作文献

Database Establishing and Data Mining of Pulmonary Diseases Based on Clinical Works by Modern Famous Veteran Doctors of Chinese Medicine CAI Yong-min, CHEN Li-ping, LI Jian-sheng, LI Qing-lei, SUN Shu-ming, and LI Cheng-wen *Basic Medical School, Henan College of Traditional Chinese Medicine, Zhengzhou (450046)*

ABSTRACT Objective To explore syndrome and treatment laws for treating diseases of the pulmonary system by establishing database based on clinical works by modern famous veteran doctors of Chinese medicine (CM). **Methods** Clinical experience and literature of medical records in clinical works by modern famous veteran doctors of CM were taken as data source. Database was established by fields and program design. On these bases, data mining methods such as frequency analysis, cluster analysis, factor analysis, and correlation laws were performed in syndrome and treatment laws for treating diseases of the pulmonary system. **Results** Established were database capable of literature searching, information statistics, data mining of modern famous veteran doctors of CM. A total of 34 414 data were input, including medical records and notes 28 045 items (81.49%) and clinical experience 6 369 items (18.51%). In medical records and notes, there were 14 048 items (50.09%) in male and 9 466 items (33.75%) in female, and the ratio of male to female was 1.48:1. There were 4 531 items (16.16%) with no marked gender in medical records or notes. Data mining such as correlation analysis, cluster analysis, factor analysis, correlation laws in more fields could be realized. **Conclusions** Medical records and notes were dominated in data collected in this paper. The prevalence of pulmonary diseases was obviously higher in males than in females. The trend of concentrated manifestations in related fields for pulmonary diseases could be surfed by this database. Diagnosis and treatment laws for treating diseases of the pulmonary system could be found by various adaptive data mining targeting different fields. Multi-variables of symptoms, syndromes, prescriptions, and herbal drugs could be data mined in large samples of clinical literatures.

基金项目:国家中医药管理局国家中医临床研究基地业务建设科研专项(No.JDZX2012027)

作者单位:河南中医学院基础医学院(郑州 450046)

通讯作者:李建生, Tel: 0371-65676568, E-mail: li_js8@163.com

DOI: 10.7661/CJIM.2015.10.1186

KEYWORDS diseases of the pulmonary system; famous veteran doctors of Chinese medicine; data mining; literature of clinical works

数据挖掘是现代研究名老中医辨证用药思路的方法之一,名老中医临床著作文献中记载了可信度较高的临床信息数据,其中载录的临证经验和医案大多为名老中医真实的诊疗记录。为寻找名老中医临床诊治肺系疾病的有关规律,本试验依托国家中医临床研究基地建设项目,以名老中医临床著作文献为采集对象,采用现代数据挖掘的方法,对肺系疾病的证治规律进行研究,冀为今后中医治疗肺系疾病提供参考。

资料与方法

1 数据采集

1.1 采集范围和要求 收录 1911 年—2011 年间名老中医的肺系临床经验及医案著作文献,其中名老中医选择标准为 1911 年—1990 年出版的相关中医著作收录有中医临证经验或医案的名医,或 1991 年以来由人事部、卫生部、国家中医药管理局命名的全国 1~4 批名老中医或国医大师。著作选择标准为 1911 年—2011 年出版的收录有名老中医临证经验或医案内容的著作。通过检索中国国家图书馆·馆藏目录及《新中国六十年中医图书总目》等书目数据库和工具书,并结合实际书目调研,采集到符合要求的著作共 718 种。

1.2 纳入及排除标准 纳入标准:著作的中医病证包括咳嗽、哮喘、喘病、咳喘、哮喘、肺胀、肺络张、肺痛、肺痿、肺癆、肺水、肺厥、肺衰、尘肺、悬饮、痰饮、肺积、肺痹、肺部肿块、咯血、肺热病、感冒、麻疹、温病、鼻病、咽喉病 26 种;西医病证包括咳嗽、急性支气管炎、细支气管炎、慢性阻塞性肺疾病、肺气肿、慢性支气管炎、慢性肺源性心脏病、支气管哮喘、支气管扩张、肺炎、肺鼠疫、肺脓肿、弥漫性间质性肺疾病、原发性支气管肺癌、细支气管肺泡癌、肺纤维瘤、支气管腺瘤、肺转移性肿瘤、肺不张、肺水肿、胸腔积液、胸膜炎、脓胸、气胸、肺囊性纤维化、呼吸衰竭、肺性脑病、急性呼吸窘迫综合征、白喉、肺结核病、普通感冒、流行性感冒、严重急性呼吸综合征、急性上呼吸道感染、反复呼吸道感染、扁桃体炎、鼻炎、咽炎、喉炎、鼻窦炎、鼻咽恶性肿瘤 41 种。具有诊治内容的中医医案及临床经验;或既有理论阐述又有医案或临床经验,且能按照数据库的录入原则(要求病名、症状或舌脉、方或药为必备字段)进行拆录入者。排除标准:只有理论论述而无具体诊治内容的著作。

2 数据库设计

2.1 字段设计 根据国家中医临床研究基地建设项目的整体设计,结合临床实际应用需求,首先预设字段,然后通过对名老中医经验医案原文的实际解析标引补充完善预设字段,最后根据本数据库的应用需求确定了 123 个字段,并在此基础上建立了分类体系。主要字段及其分类体系如下。

1 文献相关信息

1.1 书名

1.2 第一作者

……

1.6 文献类型

1.6.1 临床经验(打开“临床经验”模块)

1.6.2 医案医话(打开“医案医话”模块)

2 临床经验(可重复添加)

2.1 基本信息

2.1.1 诊治医生

……

2.1.2 中医病证信息

2.1.2.1 病证(名)(建标准表)

2.1.2.2 病因(建标准表)

2.1.2.3 病机(建标准表)

2.1.3 西医病名(建标准表)

2.2 诊治信息

2.2.1 证候分型(可重复添加)

2.2.1.1 证型(名)(建标准表)

2.2.1.2 症状(建标准表)

2.2.1.3 舌象

2.2.1.3.1 舌体(建标准表)

2.2.1.3.2 舌质(建标准表)

2.2.1.3.3 舌苔(建标准表)

2.2.1.4 脉象(建标准表)

2.2.1.5 病因(建标准表)

2.2.1.6 病机(建标准表)

2.2.1.7 西医检查

……

2.2.1.8 治则(建标准表)

2.2.1.9 治法(建标准表)

2.2.1.10 疗法

2.2.1.10.1 中药疗法

2.2.1.10.1.1 方剂(名)(建标准表)

2.2.1.10.1.2 组成中药信息(可重复添加)

2.2.1.10.1.2.1 中药(名)(建标准表)

2.2.1.10.1.2.2 药物剂量

2.2.1.10.1.2.3 剂量单位(建标准表)

2.2.1.10.1.2.4 炮制方法(建标准表)

2.2.1.10.1.3 药方剂型(建标准表)

2.2.1.10.1.4 制备方法

2.2.1.10.1.5 给药方式(建标准表)

2.2.1.10.1.6 随证加减(可重复添加)

2.2.1.10.2 西药疗法(可重复添加)

.....

3 医案医话(可重复添加)

3.1 基本信息

3.1.1 诊治医生

.....

3.1.2 患者信息

3.1.2.1 姓名

3.1.2.2 性别

.....

3.1.2.7 初诊日期

3.1.2.8 就诊地

3.2 诊治信息(可重复添加)

3.2.1 诊次

3.2.2 日期

3.2.3 症状(建标准表)

3.2.4 舌象(同“2.2.1.3 舌象”)

3.2.5 脉象(建标准表)

3.2.6 西医检查(同“2.2.1.7 西医检查”)

3.2.7 病因(建标准表)

3.2.8 病机(建标准表)

3.2.9 病证(中医病名)(建标准表)

3.2.10 西医病名(建标准表)

3.2.11 治则(建标准表)

3.2.12 治法(建标准表)

3.2.13 疗法(同“2.2.1.10 疗法”)

3.2.14 疗程

3.2.15 禁忌

3.2.16 治疗结果

3.2.16.1 日期

3.2.16.2 症状舌脉变化

3.2.16.3 西医检查变化

3.2.16.4 疗效

3.2.17 注意事项

2.2 程序设计 在以上基础上,设计具有分类解析录入、与其他相关数据库跨库检索、多条件复合统计分析、有针对性进行数据挖掘等功能的结构化数据库。具体包括标准词库、后台管理、界面录入、检索统计分析和数据挖掘等模块。其中录入界面分两栏,右栏显示原始文献的电子文档,备录入复制选用,左栏显示上述各个字段;管理界面按查询、修改、添加、删除、导入、导出、备份、恢复、帮助等功能设计;分析界面按多字段相互关联分析要求设计。

3 建标准词库 对于部分词语较多的字段,如病因、证型、症状、舌质、舌苔、脉象、治法、方剂、中药等,参照国家标准《中医临床诊疗术语》^[1]、《中医基础理论术语》^[2],国家名词委《中医药学名词》^[3]、《医学名词》^[4]及《中国中医药学主题词表》^[5]等规范类文献,建立名老中医肺病数据库标准词库。对于在录入过程中发现的标准词库中没有的词语,则按原文录入,审核人员根据原文录入的词语出现的频率及标准词的要求,经确认后可纳入标准词库。从而使用词规范统一,检索时不会出现漏检或其他错误。本研究最终对123个字段中的62个字段建立了标准词库。

4 数据录入与清洗 制定录入原则和方法,给设计的每一字段均赋予一明确定义,并标明具体录入操作方法,在此基础上对符合纳入标准的数据进行录入。对于在录入过程中出现的不完整的数据、错误的数据和重复的数据等,在录入全部完成后进行清洗,予以纠正。

5 数据挖掘 根据对象和目的不同,数据挖掘采用不同模式。采用SPSS 19.0统计软件,分别选取不同肺系疾病的高频症状、药物进行聚类、因子分析;用SAS 9.1软件对高频症状与药物间进行关联规则研究。

5.1 频数分析 频数分析用于症、证、药、方等变量出现的频数统计。

5.2 聚类分析 聚类分析用物以类聚的原理,通过对不同变量间相似程度的分析,使相似程度大的变量聚成一类,反之聚成另一类,形成由近至远,由小到大的分类系统,有助于辨识数据中的规则,用于研究症、药等变量自身间的远近关系^[6]。药物功效分布按《中药学》^[7]教材的分类表归类统计,对于教材中未收录药物的性味归经参照《中药大辞典》^[8]补充完整。

5.3 因子分析 因子分析从分析多个原始指标的相关关系入手,找到支配这种相关关系的有限个不可测量的潜在变量,并用这些潜在变量来解释原始指标之间的相关性或协方差关系,用于研究隐藏在药物和症状背后的潜在变量^[9]。

5.4 关联规则 关联规则从大量记录中发掘数据中项集之间有趣的关联或相关联系,探讨两个或多个变量的取值间存在的某种规律性,用于研究常用药物组合和方剂配伍规律^[10]。

结 果

1 建立肺病数据库 按照以上研究方案,建立了现代名老中医临床著作肺病数据库。本数据库共录入34 414条数据,其中医案医话28 045条,占

81.49%, 临床经验 6 369 条, 占 18.51%; 医案医话中男性 14 048 条, 占 50.09%, 女性 9 466 条, 占 33.75%, 男女之比为 1.48:1, 未注名性别者(均为医话文献)4 531 条, 占 16.16%。数据库具备文献检索、信息统计、数据挖掘等功能, 可以实现任意字段间的相互关联分析。

2 频数分析(表 1) 从数据库的查询平台中可直接实现频数统计, 其他所有肺系疾病的各个字段也可采用类似的方法进行频数分析(以中医病名感冒的证型、症状为例, 表 1 可见频数、频率前 10 位)。

表 1 感冒证候分型、症状 (前 10 位)

证型	频数	频率(%)	症状	频数	频率(%)
风寒袭肺证	333	18.70	咳嗽	1 716	19.47
风热犯肺证	291	16.34	发热	1 654	18.77
暑湿伤肺证	188	10.56	头痛	1 464	16.61
气虚证	167	9.38	鼻塞	893	10.13
阴虚证	152	8.53	恶寒	781	8.86
风邪袭肺证	144	8.09	流涕	646	7.33
阳虚证	138	7.75	恶寒发热	448	5.08
表虚证	124	7.02	咽痛	445	5.05
痰浊阻肺证	123	6.91	无汗	437	4.96
痰热蕴肺证	121	6.79	口渴	328	3.72

3 关联分析(表 2) 在最后确定的 123 个字段中, 任意字段间均可实现相互关联分析。以支气管哮喘为例, 采用频数分析的方法确定前 2 种高频证型, 且分别与每一证型相关的症状、舌质、舌苔、脉象、方剂、中药 6 个字段关联后可见频数较高的相关内容。采取相同的方法可进一步找出症状与药物等之间的关联。

4 聚类分析(图 1) 以高频次慢性支气管炎用药为例进行聚类分析, 形成 5 个比较有意义的聚类组合: C1 止咳平喘药: 苦杏仁、紫苏子、紫菀、款冬花、枇杷叶、桑白皮、葶苈子; C2 补气药: 甘草、大枣、太子参、党参、山药、白术、黄芪; C3 敛肺涩肠药: 五味子、五倍子、罂粟壳、诃子、肉豆蔻; C4 发散风寒药: 生姜、桂枝、紫苏、白芷、麻黄、防风、羌活; C5 清化痰热药: 川贝母、浙贝母、瓜蒌、桔梗、竹茹、天竺黄、海蛤壳。

5 因子分析(表 3) 以肺痿病为例对其高频症状进行因子分析共提取到 12 个公因子, KMO 值为 0.618, 累计贡献率 67.452%, 可见 F1 示肾不纳气, F2、F3 示心脾两虚; F4、F5、F12 示痰浊阻肺, 病性为虚寒; F6、F7、F8、F11 示阴虚肺热, 病性为虚热。

表 2 支气管哮喘前 2 种高频证型相关字段关联分析 (频数)

证型	症状	舌苔	舌质	脉象	方剂	中药
风寒袭肺证	气喘(48)	白苔(42)	淡白舌(9)	浮脉(22)	小青龙汤(23)	麻黄(65)
	咳嗽(39)	薄苔(18)	淡红舌(4)	紧脉(18)	射干麻黄汤(10)	细辛(48)
	哮鸣(37)	滑苔(18)	淡舌(3)	弱脉(7)	二陈汤(6)	甘草(43)
	恶寒(36)	润苔(6)	暗红舌(1)	细脉(2)	麻附细辛汤(4)	半夏(42)
	无汗(21)				麻黄汤(3)	五味子(40)
痰热壅肺证	咳嗽(45)	黄苔(40)	红舌(32)	数脉(37)	麻杏石甘汤(24)	杏仁(64)
	气喘(44)	腻苔(22)		滑脉(29)	清肺定喘汤(8)	麻黄(59)
	咯痰(30)	白苔(8)		弦脉(4)	清肺化痰饮(5)	甘草(40)
	喉中痰鸣(21)	厚苔(8)			葶苈大枣泻肺汤(4)	黄芩(36)
	胸闷(19)	薄苔(3)			泻白散(2)	生石膏(28)

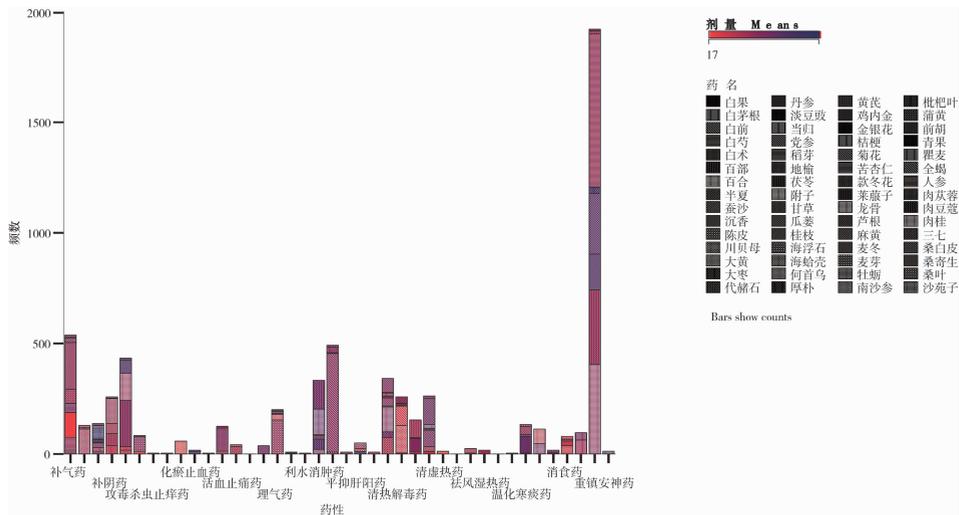


图 1 慢性支气管炎高频药物聚类分析

表 3 肺痿病医案 12 组有意义的因子载荷

因子类型	症状及因子载荷
F1	口唇紫(0.78); 动则喘甚(0.76); 心悸(0.74); 头晕(0.54)
F2	胸闷(0.77); 气喘(0.61); 气短(0.55); 呼吸困难(0.50)
F3	神疲(0.80); 倦怠(0.67); 纳差(0.62); 乏力(0.49)
F4	痰白(0.72); 痰量多(0.70); 痰有泡沫(0.64)
F5	咯痰(0.65); 咳嗽(0.65); 自汗(0.52); 痰稀薄(0.46)
F6	咽干(0.79); 口干(0.55)
F7	痰黄(0.71); 痰黏稠(0.54); 痰中带血(0.46); 盗汗(0.42)
F8	干咳(0.79); 痰量少(0.61)
F9	形体消瘦(0.82); 纳少(0.77)
F10	胸痛(0.72)
F11	气促(0.72); 口渴(0.66); 气急(0.53)
F12	肢冷(0.80); 面色萎黄(0.74)

6 关联规则(表 4) 通过对二陈汤的高频症状、药物进行关联规则分析,定义最小支持度为 2%,置信度为 70%,可见纳差、咯痰、气喘与苦杏仁关系密切,咯痰不爽、胸痞、头晕与紫苏子、白芥子、莱菔子关系密切。

讨 论

1 对数据的来源分析

寻找中医诊治疾病的规律,需要大量客观真实准确的临床信息数据,数据来源可靠是保证统计结果准确的基础。由于载体不同,或同一载体但登载信息目的不同,临床信息数据可信度差异较大。现代名老中医治疗肺系疾病积累了十分宝贵的诊疗经验,本文数据来自于记载现代名老中医诊疗经验的两类著作,一为 1911—1990 年出版的相关中医著作,二为 1991—2011 年出版的收录有全国 4 批名老中医或国医大师临证经验或医案内容的著作,同时本文尚对从上述两类著作中采集到的数据制定了明确的纳入及排除标准,力求使数据来源真实可靠,研究方法客观准确,冀

能发现名老中医诊疗肺系疾病的本质规律。尽管如此,本研究方法仍有许多待完善之处,如在制定录入原则时为尊重原著原义,允许录入者增加原著中的原词语作为部分自由词,故病因、证型、症状等字段出现了含义相同而分别录入的情况,对此在数据清洗规范时进行了合并处理,如将痰浊阻肺证、痰湿阻肺证合并为痰浊阻肺证,风热犯肺证、风热袭肺证合并为风热犯肺证等。

2 对数据的初步挖掘分析

频数分析是进行聚类、因子、关联规则等分析的基础,本数据库查询平台可直接找出肺系疾病不同字段的高频数据,如表 1 感冒的高频症状和证型,然后在此基础上针对原始数据继续用其它方法进行数据挖掘。

2.1 症-证的关系 症状是诊断疾病最直接的依据,它通过望闻问切收集第一手临床资料,中医强调治病求本,需对四诊资料正确分析,准确辨证,明确病因病机,才能随证立法遣药,制定有效方剂治疗疾病^[11]。因此准确把握症、证的关系,就能掌握中医临床思维的方法和技能。对于此类关系的挖掘可采用聚类、因子等综合分析,找出每种病的主症、次症、兼症、舌脉等。对证候可据症状推断,能因子分析得出,如表 3 中 F2、F3 的胸闷、气短、气喘、神疲、倦怠、纳差等示心脾两虚; F4、F5、F12 的痰白、痰量多、咯痰、咳嗽、痰有泡沫、痰稀薄、肢冷等示痰浊阻肺,病性虚寒; F6、F7、F8、F11 的咽干、口干、痰黄、痰黏稠、痰中带血、盗汗、干咳、痰量少、形体消瘦、纳少、口渴等示阴虚肺热,病性虚热,可知肺痿的证型主要有心脾两虚证、寒痰阻肺证、阴虚肺热证。

2.2 方-药的关系 方是药物的组合,方药的配伍规律的研究可以从以下几个方面进行,中药药性理论(四气五味与归经)与方剂配伍关系,同类方剂配

表 4 二陈汤高频症状、药物关联规则分析

关联数目	置信度 1	置信度 2	支持度	计数	症状、药物关联
3	51.77	100.00	2.21	5	头痛、恶寒 == > 苦杏仁
3	51.77	87.50	3.10	7	纳差、咯痰 == > 苦杏仁
3	51.77	83.33	2.21	5	气喘、喉中哮喘 == > 苦杏仁
3	10.18	75.00	2.65	6	咯痰不爽 == > 紫苏子、白芥子
4	51.77	100.00	3.10	7	神疲、纳差、发热 == > 苦杏仁
4	51.77	100.00	2.21	5	气喘、纳差、口渴 == > 苦杏仁
4	19.03	83.33	2.21	5	小便黄、纳差、口渴 == > 枳实
4	8.85	71.43	2.21	5	痰不易咯出、纳差 == > 苦杏仁、白术
5	51.77	100.00	2.21	5	小便黄、气喘、纳差、口渴 == > 苦杏仁
5	11.50	83.33	2.21	5	小便黄、纳差、口渴 == > 枳实、苦杏仁
5	2.21	83.33	2.21	5	胸痞、头晕 == > 紫苏子、天南星、莱菔子
5	2.21	83.33	2.21	5	胸痞、头晕 == > 紫苏子、山药、莱菔子
5	2.21	83.33	2.21	5	胸痞、头晕 == > 紫苏子、山药、苦杏仁

伍的一般规律,药物配伍的基本内容和原则,特殊配伍规律^[12]。对于以上的研究最常用因子分析、关联规则等,找出常用药物集群,分析出方剂的配伍规律。图 1 中名老中医治疗慢性支气管炎使用高频药物的聚类分析结果能看出 C1 常用的方有麻杏石甘汤、定喘汤等; C2 常见参苓白术散、四君子汤等培土生金的方及补肺气的玉屏风散; C3 常见于牡蛎散、九仙散等; C4 常用的方有止嗽散、桂枝汤等; C5 常用于清气化痰丸、贝母瓜蒌散等。

2.3 症-药的关系 症-药关系在临床中常见的是单药或药对和某一特殊症状的对应,治疗疾病时,常在辨证论治的基础上,加以对症药物,针对性的增强或改善某些突出症状的功效。如治咳嗽,善用炙紫菀、炙款冬花以对症,治疼痛类症状,喜用芍药、甘草缓急止痛以对症。对于症-药关系的挖掘亦可采取聚类、因子、关联等诸多挖掘方法。表 4 中有痰鸣、咯痰不爽、喉中哮鸣等以痰为主要诱因的症状时常用白芥子、莱菔子、紫苏子药组,且加枳实行气化痰,以助药力,此类症状和药物关联置信度最高 75%,支持度 2.65%。

2.4 症-证-方-药的关系 在本数据库中,可通过多项字段的相互关联分析找出肺系疾病症-证-方-药之间的关系,主要是通过对 123 个字段的分层检索实现的,如病证(咳嗽)-证型(痰热壅肺)-方剂(麻杏石甘汤)-药物(黄芩),表 2 提示支气管哮喘风寒袭肺证的常见症状有气喘、咳嗽、痰鸣、恶寒、无汗,以薄白苔、淡白舌或淡红舌、浮紧脉为主,常用方为小青龙汤、射干麻黄汤、二陈汤等,常用药为麻黄、细辛、甘草、半夏、五味子等。这种通过数据库字段的分层而形成的不同字段间的联系是关联分析,与关联规则数据挖掘方法不同。

在数据挖掘中,针对不同的字段可以采取其所适应的方法,有时针对同一字段也可以从不同角度将几种方法组合,如对病因、病机、证候等用频数分析,对药物用频数、聚类、因子等综合分析,对证候、方剂用频数、关联分析等,还有相应分析、回归分析、人工神经网络、贝叶斯网络等数据挖掘的方法,至于究竟采取什么方法进行挖掘与数据自身的特点和研究者的目的相关。本研究可针对某一具体的肺系疾病查寻其相关的高频病因病机、证候分型、症状、舌质、舌苔、脉象及方药,并针对不同字段进行各种适应性的数据挖掘,从而探寻各种肺系疾病的诊治规律,为科研和临床提供参考。需要说明的是,本研究在统计中先关注了一些字段相关的高频信息,这些反映了临床上大多医家普遍

的用药规律,而对频次少的变量须结合专业知识和具体的病案进行分析,这些低频的数据除了跟样本量的大小有关外,有时可能正好反映了名老中医用药的独特思路,对此有待进一步挖掘研究。此外,名老中医经验不仅体现在著作中,也见诸有关杂志等文献中,对此怎样甄别并采集其中客观真实的信息数据,以补充完善本数据库,亦待进一步研究。

总之,本研究数据来源于现代相关临床著作,其中载录了可信度较高的名老中医诊疗肺系疾病的宝贵经验;通过本项目的研究,可查寻各种肺系疾病相关字段的集中表现趋势;针对不同字段进行各种适应性的数据挖掘,能发现各种肺系疾病的诊治规律;本研究方法可实现大样本临床文献的症、证、方、药等多变量多角度的数据挖掘,对大样本的相关临床文献研究有一定借鉴意义。

参 考 文 献

- [1] 国家技术监督局. 中医临床诊疗术语: 疾病部分[M]. 北京: 中国标准出版社, 1997: 9-10.
- [2] 国家质量监督检验检疫总局. 中医基础理论术语[M]. 北京: 中国标准出版社, 2006: 44-81.
- [3] 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: 37-247.
- [4] 医学名词审定委员会. 医学名词: 第一分册[M]. 北京: 科学出版社, 1989: 47-56.
- [5] 吴兰成主编. 中国中医药学主题词表[M]. 北京: 中医古籍出版社, 2008: II 1-II 1364.
- [6] 卞雅莉. 基于数据挖掘技术的中医医案研究方法探讨[J]. 光明中医, 2011, 26(7): 1340-1341.
- [7] 高学敏主编. 新世纪全国高等中医药院校规划教材中药学[M]. 北京: 中国中医药出版社, 2002: 51-517.
- [8] 南京中医药大学主编. 中药大辞典[M]. 上海: 上海科技出版社, 2006: 13-3871.
- [9] Jae-On Kim, Charles W Mueller. Introduction to factor analysis: what it is and how to do it [M]. Beverly Hills, Calif: Sage Publications, 1978: 237-238.
- [10] 叶亮, 范欣生, 卞雅莉, 等. 古今治疗痛经的四物汤类方关联规则研究[J]. 南京中医药大学学报, 2008, 24(2): 94-96.
- [11] 江泳, 陈建杉, 江瑞云, 等. 论辨证论治的完整体系: 人-症-病-证[J]. 中医杂志, 2011, 52(17): 1447-1450.
- [12] 于友华, 王永炎, 赵宜军, 等. 方剂配伍规律的研究[J]. 中国中药杂志, 2001, 26(4): 219-221.

(收稿: 2014-09-13 修回: 2015-07-20)